Sub-Terahertz Circular Dichroism Spectroscopy of Solvated Proteins via Vector Network Analysis and Time-Domain Spectroscopy

by

Jhih-Hong Cheng

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Electronic Engineering and Computer Science

Queen Mary University of London

United Kingdom

2020

Declaration of Originality

I, Jhih-Hong Cheng, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for a degree by this or any other university.

Signature: Jhih-Hong Cheng

Date: 21st April 2020

Acknowledgements

This thesis would not have been possible without the support of the follwing amazing people. First and foremost, I'd like to express my sincere gratitude to my supervisor Dr. Robert Donnan, for guiding and helping me overcome obstacles throughout the process. His constant encouragement, generous support, and timely replies to my requests have been most heartening. The positivity and cordial atmosphere at all of our meetings helped me bring to fruition this thesis. As a supervisor, he selflessly shares his professional knowledge and has always been most patient in our communication. It brings me tremendous joy during my Ph.D. life under his supervision.

I would also like to thank my secondary supervisor Dr. Akram Alomainy and independent assessor Prof. Xiaodong Chen for giving their time and providing valuable feedbacks at each stage. If this thesis were to reflect much of the academic rigor, I owe it to them.

I also feel pleasantly obliged to acknowledge Dr. Oleksandr Sushko for dispensing his knowledge of Quasi-Optical system and THz TDS. QO system and THz TDS are important for my experimental studies. He also squeezed out time to patiently discuss and advise my first journal paper and shared his knowledge of data processing on THz TDS and protein hydration.

Furthermore, the assistance of our lab manager Dr. Max Munoz Torrico for purchasing samples and doing his best in promptly fixing equipment is also well appreciated.

I would like to extend my appreciation to Robert Jones, a Ph.D. student in our group, for his assistance on Quasi-optical system and THz TDS alignment. His advice and all of our fruitful discussions contributed to my first journal paper. We had a wonderful time in the lab and I never got bored with him during the experiment.

I would like to thank Dr. Bin Yang for allowing me access to his THz TDS and for giving his advice on Quasi-optical system and THz TDS measurement.

Many thanks to my former advisors Dr. Heather Song and Dr. Scott Trimboli at the University of Colorado, Colorado Springs for their encouragement. They helped me steady my staggering steps as I was fumbling in some soul-trying period..

Last but not least, I want to thank to my father Shiang-Lin Cheng and my mother Shih-Chieh Wang for supporting me in this never-ending odyssey of academic pursuit. Without their financial support, encouraging words and consultation, I would never achieve this success. Their love and unfathomable tolerance have sustained me throughout the 10 years of studying abroad, making me who I am now. They always have my back without saying a word or asking anything in return. I dived headlong into the research, and they went with me. To my brother Jing-Kai Weng, for patiently proofreading all of my thesis word by word. To my beloved baby sister Wei-Yun Cheng, for her support and encouragement. To Hsuan Yang, I never got bored with her company during the measurement and writing. All the completion of this undertaking could not have been possible without their contribution and participation. It gives me tremendous pride and joy to acknowledge these people and everyone that knowingly or unwittingly helped me along the way. Thank you all.

Abstract

The study of solvated proteins by terahertz (THz) frequency electromagnetic radiation, characterized by submillimeter wavelength signal beams, suffers from strong absorption by water. Yet, water is vital to normal protein function; hydration plays a formative role in a protein acquiring its operational molecular structure. To decouple the traditional conflation of water spectra with that contributed by the protein in conventional transmission or absorption spectroscopy, this thesis innovates the experimental methodology of circular dichroism (CD) spectroscopy; water being achiral. Dealing with the heavy signal loss from absorption by water is countered by the gain afforded by using a state-of-the-art vector network analyser (VNA), that is interfaced with a quasi-optical (QO) circuit, containing the sample of dissolved protein. This apparatus has been employed to co-study CD protein conformational changes conventionally acquired in the far-UV at the discrete frequency of 222 nm. This thesis explores the first acquisition of a continuous THz CD spectra of a protein as it responds to variation in the temperature, salinity and pH of its host aqueous medium, in order to lay foundations for interpreting associated spectra in relation to the given conformational state the protein adopts in response. Experimental work in this thesis was performed over the mm and sub-mm waveguide bands: WR7 (110 – 170 GHz) and WR3 (220 – 325 GHz), respectively; and, THz Time-Domain Spectroscopy (TDS). (200 – 1400 GHz). A QO-based, sub-THz CD spectrometer is demonstrated on the sample protein, myoglobin, over the waveguide bands of WR7 and WR3. In so doing, an equivalent methodology is propounded for acquiring cross-polar spectra that eliminates the need for having to physically rotate the receive-horn through 90˚. A significant source of systematic-error is thereby eliminated (and, no less, random-errors, that may occur when cabling is moved). The ‘zero-crossing’ points of CD spectra are characterized as being correlated with traditional far-UV CD spectra in the given environmental parameter settings of, for example, pH and/or sample concentration level. For the extending application, the sub-THz CD QO system studies

three solvated proteins namely, Myoglobin, Bovine Serum Albumins, and Papain at WR3. Water is used as a buffer, since it contributes no CD signal. Secondary structures of the proteins consist of different amounts of α-helices and β-sheet forms. Different optical path-lengths and concentration levels are considered in order to compare their efficiency in providing sub-THz CD signatures by different structures of proteins. The application of THz TDS is to study the dielectric properties of the three proteins. Dielectric properties of solvated proteins are strongly influenced by the flexibility of protein secondary structures, hydration bonds, and water-protein interaction. By analyzing the protein's dielectric behaviour, explanation of protein CD behaviour will be sought for the sub-THz frequency domain.

Journal publications from this thesis

[1] J. Cheng, R. Jones, O. Sushko, Y. Tashiro and R. Donnan, "Quasi-Optical Sub-THz Circular Dichroism Spectroscopy of solvated Myoglobin", *IEEE Transactions on Terahertz Science and Technology*, pp. 1-1, 2020.

Conference presentation from this thesis

[1] J. Cheng, R. Jones, O. Sushko, B. Yang, Y. Tashiro and R. Donnan, “Application of sub-THz Circular Dichroism Quasi-Optics to Probe the Conformality of Solvated Protein”, IRMMW-THz Paris, 2019 (poster)

List of Abbreviations

BSA	**Bovine Serum Albumin**
CD	**Circular Dichroism**
EOS	**Electro-Optical Sampling**
FFT	**Fast Fourier Transform**
FP	**Fabry-Perot**
MB	**Myoglobin**
PCA	**Photoconductive Antenna**
QO	**Quasi-Optic**
TDS	**Time-Doman Spectroscopy**
TF	**Transfer Function**
THz	**Terahertz**
TPX	**Poly-4-Methyl Pentene-1**
UV	**Ultraviolet**
VCD	**Vibrational Circular Dichroism**
VNA	**Vector Network Analyzer**

List of Figure

List of Table

Table of Contents

Chapter 1

Introduction

1.1 The Electromagnetic spectrum in the THz domain

The term Terahertz (THz) first occurred in the mid1970s [1]. The location of the THz band is between the microwave and infrared bands of the electromagnetic spectrum as shown in Fig. 1.1. Its domain is bounded between 0.1 to 10 THz with corresponding wavelengths between 3 to 0.03 mm [2, 3]. THz radiation is non-ionizing due to its low photon energy, and such a characteristic makes THz radiation of medical utility [4- 6]. Since its photon energy is low (meV) relative to optical frequencies and higher, it does not ionize atoms and biological molecules such as proteins [7, 8]; thus this process causes negligible-to-no radiation-damage. THz radiation penetrates nonpolar and insulating materials such as cloth, plastic, etc. while leaving the material unaltered. However, it can be absorbed by polar molecules such as water or be strongly reflected by metals [4].

The frequency domain of the THz electromagnetic spectrum, as shown in Fig. 1.2 [7], corresponds to the lower end of the far-infrared region. This is critical for the rich variety of chemical and physical processes it can probe, such as: determining the carrier-carrier lifetime in semiconductors [10]; detecting phonon vibrations in crystalline solids [11]; analysing hydration dynamics of biological molecules [12]; and, recording hydrogen bond fluctuation in solvents [13]. However, the application of THz technology has its fair share of limitations. THz radiation is easily absorbed by high-water-content samples such as blood and biological cells. Most research in recent years has been focusing on the

variation of absorption coefficient with different concentration of protein solutions [10-15], or structural alteration at different temperatures [12, 14-17]. Studies on the influence of the hydration layer surrounding the proteins, or on dynamic hydrogen bonding in liquids, have also increased significantly in recent years [12, 15, 18].

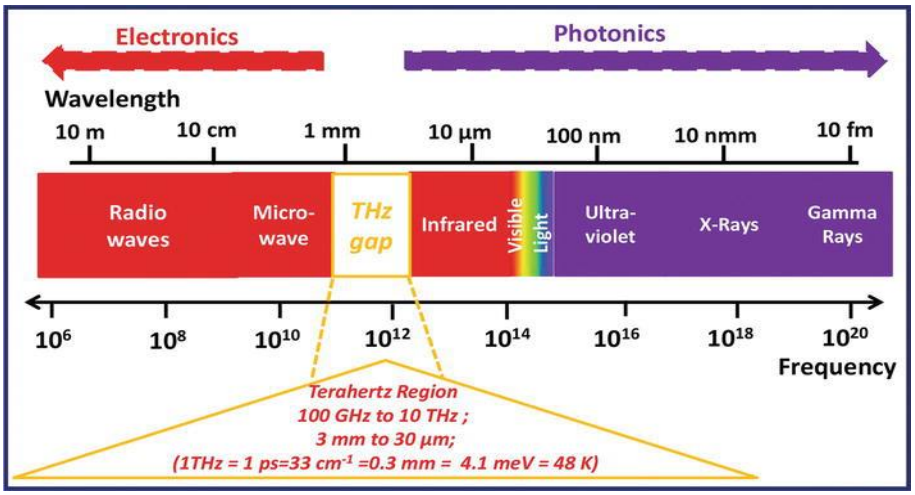

Figure 1.1: The span of the THz band of the electromagnetic spectrum [2]. It covers from 100 GHz to 10 THz. The so-called 'THz gap' is located between microwave and infrared bands and is shrinking with advancing technology. The low photon energy (meV) of THz radiation does not ionize atoms and molecules, making it suitable for biological study.

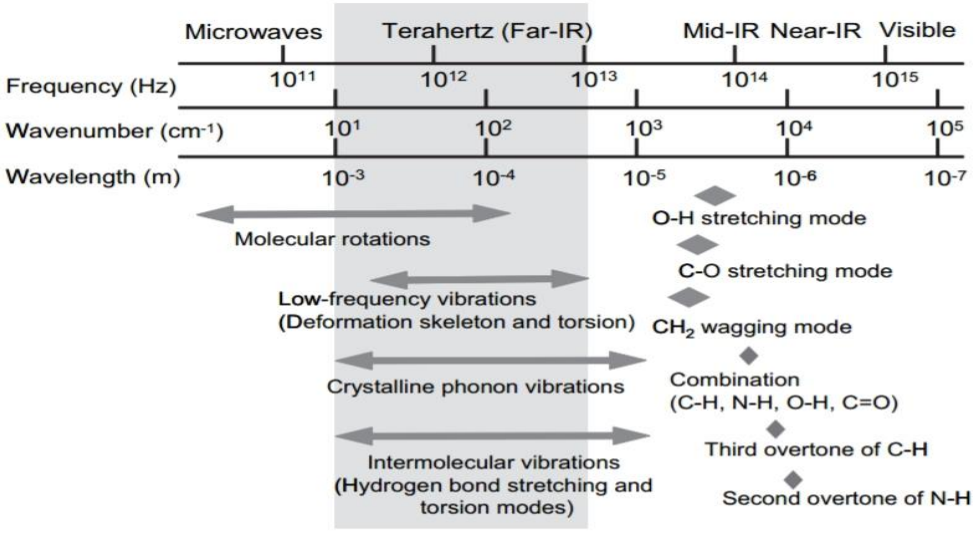

Figure 1.2: Characteristic infrared absorption frequencies of chemical compounds [7]. THz radiation is sensitive to various molecular modes of oscillation. Molecular rotational and vibrational modes can be detected by THz waves and influence their absorption. Hydrogen bonds play an important role in stabilizing the structures of proteins, and such water-protein interactions result in different THz circular dichroism behaviour and dielectric properties. Such characteristics make THz radiation suitable for identifying and analysing solvated proteins.

1.2 Application of Terahertz frequency electromagnetic radiation to biomolecular study

For decades, THz electromagnetic (EM) radiation has been widely used to study biological samples from amino acids, through to cells, upto the tissue level. Compared to X-ray crystallography, nuclear magnetic resonance spectroscopy, and microwave dielectric spectroscopy, THz EM radiation has a lower rate of scattering than infrared and visible light in analysing biological tissue, especially of blood inside the blood vessel [19]. With a reliable and highly sensitive optical system using femtosecond laser technology, THz spectroscopy such as THz Time-Domain Spectroscopy (THz-TDS) has provided significant complementary information about the structure and function of proteins [12, 14, 20, 21] and organic tissue [4, 6, 7, 22, 23], when operating in an imaging mode [5, 24]. Furthermore, millimeter- and submillimeter EM radiation has a spatio-temporal scale that efficiently couples to protein vibrational modes [25, 26] and dielectric water relaxation [13, 27, 28] fundamental to many-fold chemical and physical processes such as those active in the hydration dynamics of biological molecules [10, 29].

In the sub-THz spectral domain (100 to 300 GHz), the interaction of EM radiation with material of polymer form is to drive free or hindered rotational modes of molecular motion with some energy dissipated as heat. With increasing frequency photon energy rises generally in increased EM absorption as more molecular vibrational modes are coupled to. The group of Markelz was one of the earliest to demonstrate that low-frequency protein vibrational modes are optically active from 0.2 to 2 THz [12, 13, 27]. The finding indicated that molecular rotational and vibrational modes we reach in the sub-THz and THz spectral bands. THz dielectric spectroscopy and THz-TDS have been applied by the group of

Markelz to study the conformational change of proteins and to identify biomolecules by their influence of their structure. Further studies are evidence that biological study is gaining in importance at THz frequencies. Wang *et al* [31] have successfully used THz absorption spectroscopy to detect a type of intracellular metabolite, riboflavin. B. subtilis cells; the intracellular metabolite riboflavin exhibits distinctive absorption features within the frequency domain 0.5 to 2.5 THz. The THz spectrum is also showing its potential in studying cancerous tissue. Such tissue is embedded in paraffin to mitigate against the influence of water [22, 23, 32]. Results show that cancerous tissue has a higher absorption coefficient relative to healthy tissue. Comparison of data between normal and tumorous gastric tissue showed clear differences at low THz frequencies [32]. Kun *et al* [23], report THz absorption coefficient curves of brain glioma and normal brain tissues to also begin overlapping near 1.4 THz. As several studies have indicated, water-rich samples, such as biological tissue, can be measured up to 1 THz.

While water is a strong absorber, it remains necessary for the proper structure of biomolecules and for enabling proteins to adopt and hold their folded, functional state [15]. Two-dimensional femtosecond time-domain spectroscopy has become an important tool for studying the dynamic and coupling in electronic, vibrational and other molecular excitations in liquids [33-36]. However, the main reason for the lack of implementation of 2D THz spectroscopy in the study of the properties of biomolecules is largely attributed to the requirement of at least several μJ of THz pulse-energy to generate enough action for triggering macroscopic polarization inside the molecule [33]. In the study of solvated proteins, the concentration level, temperature, and value of pH can significantly influence the THz absorption of proteins. One such study employed a sub-THz quasi-optical (QO)

system [37] to measure 220 to 325 GHz EM absorption of solvated proteins having concentrations ranging from 0.25 to 100 mg/ml. The absorption coefficient decreased as the concentration level rose due to the slow rotational motion. However, Sushko *et al* [37] found that the absorption coefficient is not influenced by the hydration shell enveloping the protein but rather by the existence of a large protein dipole moment. According to Markelz *et al* [13], slowing molecule rotational motion could reduce the bulk-water background absorption. Disruption of existing hydration bond networks occurred in pure liquids when foreign molecules of slower rotational dynamics were introduced, resulting in the reduction of terahertz absorption [4]. Overall absorption of mixtures decreased when the concentration level was raised. The experiments of Ebbinghaus *et al* confirmed this [17]. Temperature and pH values can influence hydration bonding inside the structure of protein. Human serum albumin (HSA) is a good example of temperature sensitivity. Luong *et al* [10] found that HSA transformed into an unfolded state with permanent disruption in its structure when temperature is raised to 70℃, leading to an irreversible change in THz absorption. Vondracek *et al* [14] revealed that the value of pH has a clear influence on the THz absorption of β-lactoglobulin; and Ebbinghaus *et al* [17], have shown that the higher the level of pH (of the protein solution), the higher the capacity for THz radiation to be absorbed. Based on the data, the THz absorption at pH 2 and pH 5 decreased linearly.

According to Singh *et al* [38], the backbone of a protein is less defined at lower hydration levels, indicating that a protein sample must be hydrated to at least 30% of its total weight. Water forms a hydrogen bonding network surrounding the protein, and the coupling between protein and hydration-water (as opposed to bulk-water), by H-bonding, triggers strong interactions [38, 39]. In early experimentation using THz spectroscopy,

biological samples were mostly prepared as dried powder precisely to avoid water absorption [40]. The application of circular dichroism (CD) at THz frequencies to the study of solvated protein is drawing increasing interest [41-44]. Since water is achiral and is not optically-active, its insensitivity to the THz CD signal is advantageous in yielding spectra free of water signatures. The first THz CD simulation in 2003 [41] applied *ab initio* analysis based on a 'mass and spring' model for the helical structure of a molecule. The resulting zero-crossing points (i.e. frequencies at which the CD value equals zero) are taken to be a THz CD 'fingerprint'. However, the secondary structure of proteins has many-fold vibrational and rotational normal modes. The simple 'mass and spring' helical model captured broadly characteristic, but not detailed CD response. In 2010, the vibrational circular dichroism (VCD) spectrum of water in the gas phase has simulated by Tina *et al* [44]. Although THz CD techniques have been used to study the chiral structure of materials such as gammadion-shaped gold [43], and of metamaterials [42, 45], it has not yet been applied to study hydrated proteins. Although the challenge to the methodology of THz CD remain, it has showed promise for studying biomolecules in their liquid state.

1.3 Thesis outlines

The work presented in this thesis focuses on new THz CD techniques in studying dissolved protein. Most of the findings are angled toward developing a sub-THz CD quasi-optical system and finding bio-signatures of aqueous protein samples varying in levels of concentration and values of pH. THz-TDS is used to study the dielectric response of protein and to analyse sub-THz CD spectra.

Chapter 2 outlines for background understanding pertinent theory of electromagnetic wave propagation in dielectric material and circular dichroism. The theory of the Jones'

Calculus in transforming linearly-polarized waves to left- and right-hand circularly-polarized waves is basic to CD analysis and will be mentioned in the last section.

Chapter 3 introduces the THz Quasi-Optical system and THz Time-Domain Spectroscopy as used in the Antenna Lab, Queen Mary University of London. Related components and protein samples will be introduced in the chapter.

Chapter 4 presents a new sub-THz CD quasi-optical (sub-THz CDQO) system in studying dissolving Myoglobin at different values of pH, and at different concentrations of buffers.

Chapter 5 presents the study of biological signatures of dissolved Myoglobin, Papain, and BSA samples with a new sub-THz CD QO circuit using different optical path-lengths and sample concentrations.

Chapter 6 concludes the work in the dielectric study of protein samples on the THz TDS and correlates them with sub-THz CD data in chapter 5.

Chapter 7 concludes the thesis and scopes area for future work.

Chapter 2

Theory of Electromagnetic wave propagation in Liquids and Circular Dichroism

2.1 Interaction of Electromagnetic Waves with Matter

2.1.1 The Maxwell Equations

The electromagnetic wave is the coupled, co-propagation of an electric and magnetic field described by Maxwell's equations. These equations are built based on Gauss' law for electric and magnetic fields, Faraday's law of induction, and Ampere's law with Maxwell's correction for continuity of charge. The initial integral form of Maxwell's equation are [46]:

$$\oint \vec{\mathrm{E}} \cdot \overrightarrow{dA} = \frac{q}{\varepsilon_0} \tag{2.1}$$

$$\oint \vec{B} \cdot \overrightarrow{dA} = 0 \tag{2.2}$$

$$\oint \vec{\mathrm{E}} \cdot \overrightarrow{ds} = -\frac{\partial}{\partial t} \oint \vec{B} \cdot \overrightarrow{dA} \tag{2.3}$$

$$\oint \vec{B} \cdot \overrightarrow{ds} = \mu_0 i + \mu_0 \varepsilon_0 \frac{\partial}{\partial t} \int \vec{\mathrm{E}} \cdot \overrightarrow{dA} \tag{2.4}$$

$\vec{E}$ is the electric field and $\vec{B}$ the magnetic field. q is the total charge. dA is the elemental surface area of integration threaded by the flux of $\vec{B}$. $\mu_0 = 4\pi \times 10^{-7}\ N.A^{-1}$ and $\varepsilon_0 \equiv \frac{1}{\mu_0 c^2} \cong 8.854\ 187\ 817 \ldots \times 10^{-12} Fm^{-1}$ are, respectively, the fundamental electric and magnetic constants. c is the speed of light in a vacuum, being exactly 299 792 485 $m.s^{-1}$. Gauss' law relates the flux of $\vec{E}$ to a source of charge, q. Gauss' law is built on Coulomb's law relating the mechanical quantities of distance and force to field properties of electric

charge. The principles of a Gauss' analysis applied to magnetism describes the observation that the net sum of magnetic flux through a notional closed surface about a source of magnetic field is zero (i.e. magnetic monopoles do not exist). Maxwell's integral equation (2.3) summarizes the experimental observations of Faraday that dynamic magnetic flux induces an electrical current. With (2.4) Maxwell summarizes the experimental observation of Ampere, but significantly, correcting it to be consistent with the condition for continuity of charge.

2.1.2 Electromagnetic Waves in Matter

In free space, it is assumed that the total charge and current density are zero. Taking the curl of the differential forms of (2.3) and (2.4), respectively yields the wave equation of the electric and magnetic fields giving [47]

$$\nabla^2 E = \mu_0 \varepsilon_0 \frac{\partial E}{\partial t} \tag{2.5}$$

$$\nabla^2 E = \mu_0 \varepsilon_0 \frac{\partial B}{\partial t} \tag{2.6}$$

In a free space (i.e. vacuum), the speed of light is given by $c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$. When an electric field propagates through material, the phase velocity of the electric field slows in relation to the permittivity of the material, as according to $v = \frac{1}{\sqrt{\mu_0 \mu_r \varepsilon_0 \varepsilon_r}}$. The ratio between the wave speeds in vacuum and in the material is the refractive index

$$n(v) \equiv \frac{c}{v} = \sqrt{\mu_r \varepsilon_r(v)}. \tag{2.7}$$

ε_r and μ_r are, respectively, the relative permittivity and relative permeability of the dielectric. Most materials are non-magnetic at optical frequencies and have a relative permeability near unity. Its role in delineating the optical path of propagation is given by

Snell's law (describing the relationship between the angles of incidence and reflection or refraction when light meets the interface between two materials of differing refractive index ($\frac{\sin\theta_2}{\sin\theta_1} = \frac{n_1}{n_2}$)). The propagating wave amplitude for both transmission and reflection for varying angle-of-incidence as the light impinges at the boundary between a dielectric and its host medium, is given by the Fresnel expressions (the optical analog of the S-parameters in vector network analysis). However, when an electric field pass through a material, its energy is always attenuated. This can be taken into account by defining the complex refractive index. The complex refractive index is conventionally given as:

$$\tilde{n}(v) = n(v) - ik(v) \tag{2.8}$$

The imaginary part $k(v)$ is the extinction coefficient associated with attenuation of the amplitude of light as it propagates into the material. If an electric field is propagating in the so-called 'z' direction (conventionally given to be the direction of propagation), the general plane wave solution of Eq. (2.5) is:

$$\vec{E} = E_0 e^{i(\hat{k}z - \omega t)} \tag{2.9}$$

where $\hat{k} = \frac{\omega}{v}$. The velocity of the propagating electric field in material of complex refractive index can be re-written from Eq. (2.7)

$$\frac{1}{V(v)} = \frac{n(v)}{c} - \frac{ik(v)}{c} \tag{2.10}$$

Therefore, substituting (2.10) into (2.9), the equation can be re-written [47]

$$E = E_0 e^{i\left(\frac{n(\nu)\omega z}{c} + \frac{k(\nu)\omega z}{c} - \omega t\right)} = E_0 e^{i\left(\frac{n(\nu)\omega z}{c} - \omega t\right)} e^{\left(\frac{-k(\nu)\omega z}{c}\right)} \quad (2.11)$$

The term $E_0 e^{i\left(\frac{n(\nu)\omega z}{c} - \omega t\right)}$ describes the electric field propagating in the material. The component $e^{\left(\frac{-k(\nu)\omega z}{c}\right)}$ describes exponential decay with depth of penetration depth. Since the intensity is proportional to the square of the electric field amplitude, the exponential decay is squared as well. In this case, the exponential decay can be written as

$$e^{\left(\frac{-2k(\nu)\omega z}{c}\right)} = e^{(-\alpha(\nu)z)} \quad (2.12)$$

$\alpha(\nu)$ is the absorption coefficient. It is a measure of the distance traveled by the electromagnetic wave inside the sample before reaching a value of e^{-1}. Analogous with Eq. (2.7), the interrelations for complex relative permittivity and complex refractive index are conventionally expressed as:

$$\varepsilon'(\nu) = n(\nu)^2 - k(\nu)^2$$

$$\varepsilon''(\nu) = 2n(\nu)k(\nu) \quad (2.13)$$

where $\varepsilon'(\nu)$ and $\varepsilon''(\nu)$ are the real and imaginary parts of the complex dielectric permittivity of the material. In the dielectric study of protein, dielectric properties play a critical role in understanding secondary structure. Furthermore, the water-protein interaction mediated by hydrogen bonds will influence the local electrostatic environment, leading to unique dielectric permittivities for different proteins.

2.2 Theory of Circular Dichroism

2.2.1 Chirality

Biological polymers, such as amino acids, peptides, proteins and other higher structural levels of biopolymers (DNA and RNA) that are folded into larger structures, all exhibit chirality. Like our right and left hands, stereo structure of these that can superpose as mirror images of each other are called enantiomers. Amino acids and sugars are good examples of such structures. In nature, most of living organisms are built with amino acids of left-handed chirality and sugars of right-handed [48]. However, some bacteria use D-amino acids. In the family of amino acids, so-called 'D-form' amino acids (the stereogenic carbon alpha to the amino group has the D-configuration), are the exception. These are right-handed (Fig. 2.1), and uncommon in living organisms. Left-handed L-amino acids constitute proteins imbuing them with their left-handed optical character.

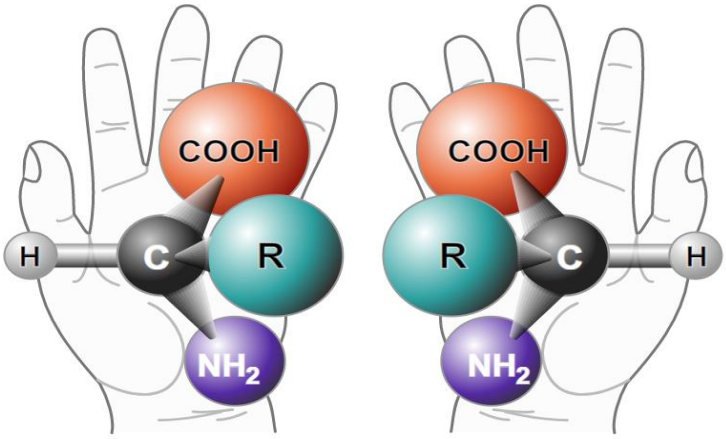

Figure 2.1: Examples of chiral structures, L- and D-amino acids [52] forming an enantiomer pair. The two stereoisomers such as L- and D-amino acids are non-superposable mirror images of each other. Amino acids are all left-handed in the living organisms.

2.2.2 Application of Circular Dichroism Measurement

Circular Dichroism (CD) is the difference in absorption between left- (LHCP) and right-handed circularly-polarized light (RHCP), and such phenomena occur in molecules containing chiral chromophores. CD spectroscopy has long-standing application in studying biomolecules with optical light, particularly in the ultra-violet (UV) portion of the electromagnetic spectrum [48]. Far UV CD spectroscopy is widely employed to study the folded and unfolded states of polypeptides and proteins [49-54]. The far UV CD spectrum (190-250 nm) of proteins reveals characteristic features of their so-called "secondary structures". The chromophore is the peptide assembly that influences the optical absorption level below 240 nm. Different types of protein secondary conformational structures (α-helix , β-sheet , and random coil) are distinguishable via their characteristic CD UV absorption 'fingerprint' [54]. Fig. 2.2 shows spectra of three conformations in the far-UV CD [56]. While CD cannot predict the numbers or locations of alpha-helixes within molecules, it enables studying secondary structure change with temperature variation or denaturation. In this case, thermodynamic information of biomolecules is made available.

For the tertiary structure of proteins, the near-UV CD spectrum provides a valuable fingerprint for identification and categorization of wild-type and mutant proteins [55, 56, 57]. For aromatic amino acids, their spectra are mostly in the near-UV CD region (260 to 320 nm). Infrared (IR) light has been employed in Vibrational Circular Dichroism (VCD) to study the 3D framework of small organic molecules (such as protein, DNA, nucleic acids, and peptides) [58, 59]. The measurement of UV CD absorption is based on the Beer's-Lambert-Bouguer law. The relationship of absorbance, concentration, and cuvette path length is written as

$$A = \varepsilon \times l \times c \tag{2.14}$$

where ε is the molar extinction coefficient, l is the path length, and c is sample concentration. The concentration level and path length of the cuvette are critical in far-UV CD measurement. Normally, for secondary structure study (180-250 nm), 0.1 mg/ml of samples are recommended for cuvettes of 1 mm path length. However, in near-UV and visible CD spectra, the signals are much weaker than in the far-UV regions. The range of protein concentration is between 0.5 to 2 mg/ml for path-lengths from 5 to 20 mm [55]. Unlike pure water, the structural features of aqueous buffer systems such as phosphate and sulfate often exhibit different absorption capacity of circularly-polarized light. To lower the strong absorption, such aqueous buffers need to be diluted to a range between 10 to 50 mM. In the far-UV CD region, Tris (tris(hydroxymethyl)aminomethane) buffer system should be completely avoided in the measurement.

THz CD techniques are only very recent in their application to the study of biomaterials and chiral materials [60, 61]. For amino acids, proteins, and other biological molecules that are chiral, or exhibiting helical secondary structure, it is reasonable to expect CD to be exhibited at THz frequencies as well as at optical. THz CD is especially sensitive to the secondary structures of proteins, enabling it to differentiate close variants in biological signatures among different biomolecules. As chiral objects, biological polymers show strong absorption to THz light by large-amplitude, collective vibrational modes [62]. Some THz vibrational CD spectroscopies have been developed to study biomolecules [60, 63] and the molecular dynamics of peptides [64].

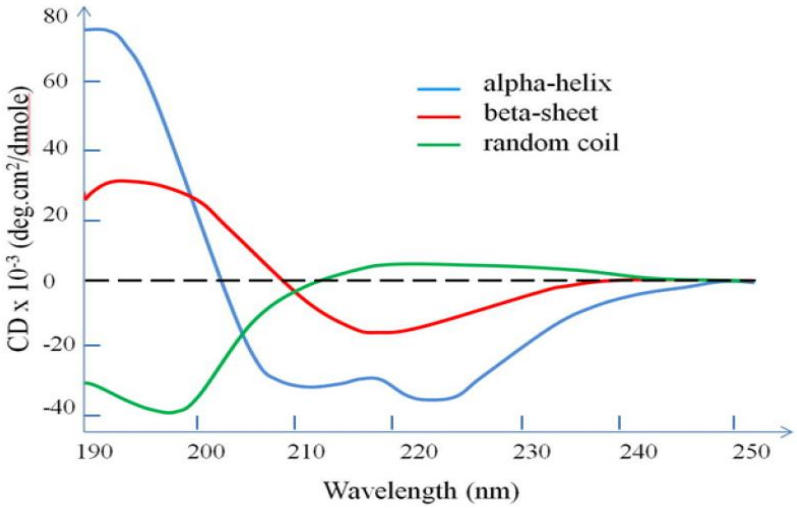

Figure 2.2: Standard Far-UV CD spectra of three secondary structures of a polypeptide chain from Y. Wei [53]. α-helix have positive bands at 192 nm and negative bands at 209 and 222 nm (blue). The bio-molecular with well-defined β-sheets have positive bands at 195 nm and negative bands at 218 nm (red). The molecular in disordered form (unfolding) has negative bands at 198 nm and very low ellipticity above 210 nm. Far-UV CD spectroscopy can be used to monitor bio-molecular (such as amino acids, peptides and proteins) conformational changes at different temperatures, pH values and predict percentage of segments of α-helices and β-sheets.

2.2.3 Circular Dichroism and the Jones' Calculus

Circular Dichroism (CD) is the difference in the absorption of left- to right-handed circularly polarized light (LHCP minus RHCP), and such phenomena occurs in molecules containing chiral chromophores. THz spectroscopy, as implemented with quasi-optical networks, conventionally operate by supporting the propagation of linearly-polarised signal beams. Left- and right-hand circularly polarised signal beams for CD spectroscopy are constructed in post-processing from the superposition of component co- and cross-polarised linear signal beams. The cross-polarised component of linear-polarisation is specially generated from the linear, co-polar signal beam via the interposing of a ferrite plate just prior to the receiving horn-antenna. The phenomenon of rotating an incoming linearly-polarised wave through a fixed angle according to the thickness of a ferrite plate, is called Faraday rotation. The Faraday rotation is caused by left and right hand circularly

polarized waves propagating at different speeds, influenced by the magnetic field in a magnetized medium. Different speeds between left- and right-hand circular waves (circular birefringence) cause the plane of the linearly polarized signal to rotate [65]. The rational and methodology for this is described by Yang *et al* [66]. It is adopted here to reduce systematic error that normally attends the acquisition of the linear cross-polarisation component attending the manual re-setting of the receiver-horn after rotating it on-axis through 90°. Eliminating the need to have to physically rotate the receiver-horn preserves phase-calibration of the quasi-optical circuit, since co-axial signal cables remain otherwise been disturbed. Fig. 2.3 depicts the progress of a linearly-polarised signal beam traversing the planar ferrite element and biological sample.

A linearly-polarised wave can be considered as the superposition of two, synchronous, (ideally) equal-amplitude, circularly-polarised oscillations but with mutually opposite senses of rotation (see Fig. 2.4) [47, 67]. Equations of the two orthogonal waves can be written as follows:

$$E_x = E_0 \cos(2\pi f_1 t + \tau) \tag{2.15}$$

$$E_y = E_0 \cos\left(2\pi f_1 t + \tau + \frac{\pi}{2}\right) \tag{2.16}$$

where E_0 is the common amplitude of oscillation and τ is its initial phase. x and y are mutually perpendicular directions. If the two orthogonal, linear, harmonic motions were to have a common initial phase τ but have a relative phase difference of $\pi/2$, then x and y components of left-handed circular oscillation are

$$E_{x-left} = E_0 \cos(2\pi f_1 t) \tag{2.17}$$

$$E_{y-left} = -E_0 \sin(2\pi f_1 t). \quad (2.18)$$

Right-hand circulation would result from either reversing time or having negative frequency such that

$$E_{x-right} = E_0 \cos(-2\pi f_1 t) \quad (2.19)$$

$$E_{y-right} = E_0 \sin(-2\pi f_1 t) \quad (2.20)$$

To obtain the resultant horizontal (x-axis) linear polarization, x-axis components of left- and right-hand circular polarizations are summed to obtain $2E_0 cos(2\pi f_1 t)$ and 0 on vertical (y-axis). Vertical linear polarization similarly follows, but by subtracting respective x- and y-components of left- and right circular polarizations; with the result $2E_0 sin(2\pi f_1 t)$ on the y-axis and 0 on x-axis [47].

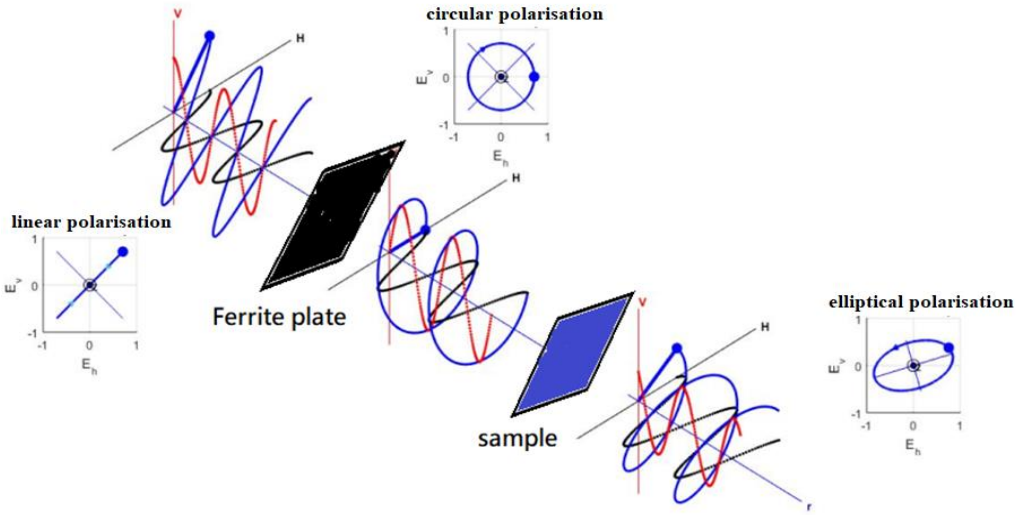

Figure 2.3: Schematic of a linearly-polarised wave traversing through a ferrite plate and the sample. The red trace represents the vertical (linear) electric field polarisation; the black trace, the horizontal (linear) polarisation; and, the blue trace the superposition of the red and black traces. The ferrite plate induces Faraday effect to rotate linearly polarized wave with certain angle (To obtain pure circularly polarized wave, the thickness of ferrite plate needs to be increased. However, such thickness of the plate can strongly absorb the power of input signal and case no signal detection after the sample). In this thesis, linearly polarized wave can be rotated 45° by ferrite plate. When circularly polarized wave penetrates through the biological sample, the different absorption of vertical (red) and horizontal (black) components can be separated by a wire-grid polariser [22]. RHCP and LHCP thereby can be calculated by Jones Calculus.

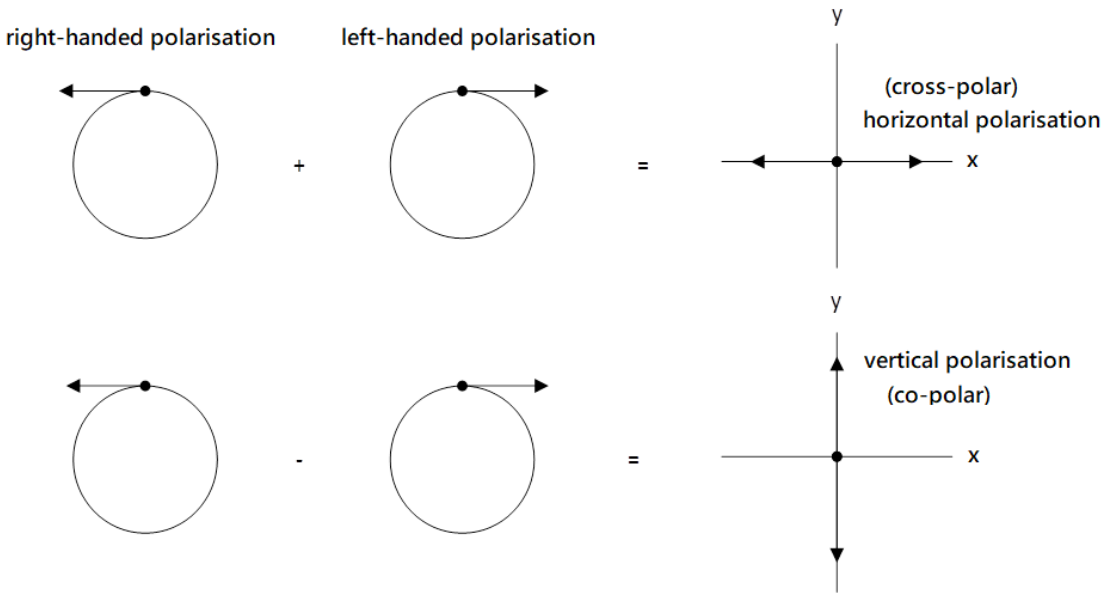

Figure 2.4: The superposition of RHCP and LHCP The sum of RHCP and LHCP waves equate to horizontal (cross-polar) linear polarisation. The difference in RHCP and LHCP waves equates to vertical (co-polar) linear polarisation. Co-polar signals are directly detected by quasi-optical circuit. Two methods can be employed to measure cross-polar signals; either by rotating the receiving horn through 90° or, by introducing a 45^o wire-grid polariser and calculating the cross-polar signal using the Jones calculus.

Two perpendicular, linear, transmission signals, co-polar (T_{co}) and cross-polar (T_{cr}), will be measured and calculated using Jones vectors. First, the matrix of the polarized wave can be represented as

$$\begin{bmatrix} E_x e^{i(2\pi ft)} \\ E_y e^{i(2\pi ft+\tau)} \end{bmatrix} = \frac{e^{i(2\pi ft)}}{\sqrt{E_x^2+E_y^2}} \begin{bmatrix} x \\ y e^{i\tau} \end{bmatrix} \tag{2.21}$$

where τ is phase difference between x and y directions and $x = \frac{E_x}{\sqrt{E_x^2+E_y^2}}$ and $y = \frac{E_y}{\sqrt{E_x^2+E_y^2}}$, respectively [66]. The Jones vector $\begin{bmatrix} x \\ y \end{bmatrix}$ for $T_{co}$and T_{cr} transmission is therefore [47, 66]

$$Tco = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$Tcr = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{2.22}$$

Since the Jones vector is by construction, normalized, (2.21) can be written as

$$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\ e^{i\tau}\end{bmatrix}. \tag{2.23}$$

For right hand and left hand circular polarizations, the values of $\tau = \frac{\pi}{2}$ and $\frac{3\pi}{2}$ are respectively substituted into the (2.23). These may be compactly put as

$$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\ \pm i\end{bmatrix}. \tag{2.24}$$

As mentioned above, the circular dichroism spectrum is calculated based on co-polar and cross-polar transmission signals measured by the THz CD QO system with Jones vectors. Co-polar (T_{co}) and cross-polar (T_{cr}) signals are respectively substituted into (2.22) and (2.24). Expressions for both right- and left- hand circularly polarized waves can be written as

$$T_{+}(\omega) \equiv (Tco(\omega) - \ iTcr(\omega))\frac{1}{\sqrt{2}}$$

$$T_{-}(\omega) \equiv \left(Tco(\omega) + \ iTcr(\omega)\right)\frac{1}{\sqrt{2}} \tag{2.25}$$

$T_{+}(\omega)$ and $T_{-}(\omega)$ respectively represent the right-hand and left-hand circularly polarized oscillations. The circular dichroism spectrum can then be calculated as

$$\Delta\mathrm{T}(\omega) = \ T_{+}(\omega) - T_{-}(\omega) \tag{2.26}$$

The formulation above only describes how to use pure linearly polarised vertical and horizontal signal beams to calculate right- and left-handed circularly polarized waves and CD spectra. In Chapter 4 and 5, a 45^{o} wire-grid will be used to detect linearly polarized

signal beams at 45^o with respect to the x-axis to reduce systematic errors in obtaining the cross-polar signal beam component. For linearly polarized signals at 45^o and -45^o, substituting the value of $\tau = 0$ and π into (2.20) gives [47]:

$$T_{45}(\omega) = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, x + \frac{\pi}{4}$$

$$T_{-45}(\omega) = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}, x - \frac{\pi}{4} \qquad (2.27)$$

By substituting (2.22) into (2.27), the equation can be rewritten as

$$T_{45}(\omega) = \frac{1}{\sqrt{2}}(T_{co}(\omega) + T_{cr}(\omega))$$

$$T_{-45}(\omega) = \frac{1}{\sqrt{2}}(T_{co}(\omega) - T_{cr}(\omega)) \qquad (2.28)$$

Table 2.1 Jones vectors

State	Vector
Vertical (co-polar) Tco	$\begin{bmatrix}0\\1\end{bmatrix}$
Horizontal (cross-polar) Tcr	$\begin{bmatrix}1\\0\end{bmatrix}$
Right-handed circular polarization, T_{right}	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\i\end{bmatrix}$
Left-handed circular polarization, T_{left}	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-i\end{bmatrix}$
45° from x axis, T_{45}	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}; \left(x + \frac{\pi}{4}\right)$
-45° from x axis, T_{-45}	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}; \left(x - \frac{\pi}{4}\right)$

Summary

In this chapter, the theory of electromagnetic wave in the matter, Circular Dichroism, and Jones' Calculus have been introduced. The equations of refractive index and jones' vectors are derived from wave equations. Since the sub-THz Circular Dichroism Quasi-optical system can only detect linearly polarised signal, Jones' Calculus is critical to convert detecting signal from linear to circular polarisation and calculate right- and left-hand circularly polarised waves. By introducing a wire-grid polarizer, the horizontal signal Tcr can be derived using Eq. (2.28). The traditional method of THz CD measurement, and the new method using a wire-grid polarizer, will be carefully explored. Since dielectric properties play a critical role in understanding secondary structure and CD behaviors of proteins are sensitive to the protein secondary structures, we are expecting to understand protein CD response by analysing protein dielectric properties.

Chapter 3

Millimeter-wave Quasi-Optical

and

THz Time-Domain Spectroscopy

3.1 The Quasi-optical Network

Quasi-Optics (QO) is the regime where the physical length dimensions of circuit (or network) components are on the order of the signal-wavelength they are carrying. Fig. 3.1 shows a photo and associated schematic of the QO network employed. It is a mm-wave (or sub-THz), network coupled to a four-port vector network analyser (VNA) (Agilent PNA-X N5244A). The baseband of this VNA is 10 MHz to 43.5 GHz. Successive transmit and receive frequency-multiplier modules (components in blue casing, see Fig. 3.1 (a)), elevate the baseband by multiplication stages to occupy V (50 to 75 GHz), W (75 to 110 GHz), D (110 to 170 GHz), and H (220 to 325 GHz) waveguide frequency bands. The circuit schematic of Fig. 3.1(c) shows a transmit horn (H1), illuminating an off-axis ellipsoidal mirror (F1). H1 is an ultra-Gaussian, corrugated horn that efficiently couples to F1. F1 collimates the beam field from H1 and illuminates the first of a confocal pair of fast, parabolic mirrors, F2 and F3. A given sample, S, is situated at their common focus. Mirror F4 and horn H2 reverse the operation of H1 and F1; the emerging collimated beam field from F3 is condensed by F4 to efficiently couple into the receive horn, H2. Ellipsoids F1 and F4 have a focal length of 250 mm; and parabolics, F2 and F3, 100 mm. The circuit measures the full set of complex-valued S-parameters in a sweep over a given waveguide

band. Post-processing transforms these raw S-parameter data into a measure of the dispersive, complex refractive index (or, permittivity), of the sample. While being comparatively narrowband in performance, compared to conventional optical spectrometers, the VNA-based QO circuit is characterized by a superior low-noise performance (0.01 dB) and higher dynamic range (> 70 dB).

Extremely low-loss transmission in free-space is the key advantage of Gaussian beam optics. Compared to custom, QMUL, THz Time-Domain spectroscopy (THz TDS), the calibration and operation of the QO circuit is simple. Accuracy in the optical alignment of circuit components is factory-set by fixed, non-adjustable, dowel-location. QO circuit layout is readily reconfigurable on optical base-plate perforated with a matrix of location-holes on a 1-inch square lattice. This flexibility allows for the introduction of further QO elements, such as wire-grid polarisers, which perform signal-conditioning operations. The VNA can support frequency-extension modules upto 1.1 THz; though current provision is only upto 325 GHz. With all these considerations, the VNA-driven QO circuit provides an easy, compact means for dispersive study of CD compared to the current-standing means available only to major facilities, such as the Free Electron Laser (FEL) at UCSB [68], which at best only offered CD measurements at ‘spot’ frequencies, not continuous, as enabled by the VNA.

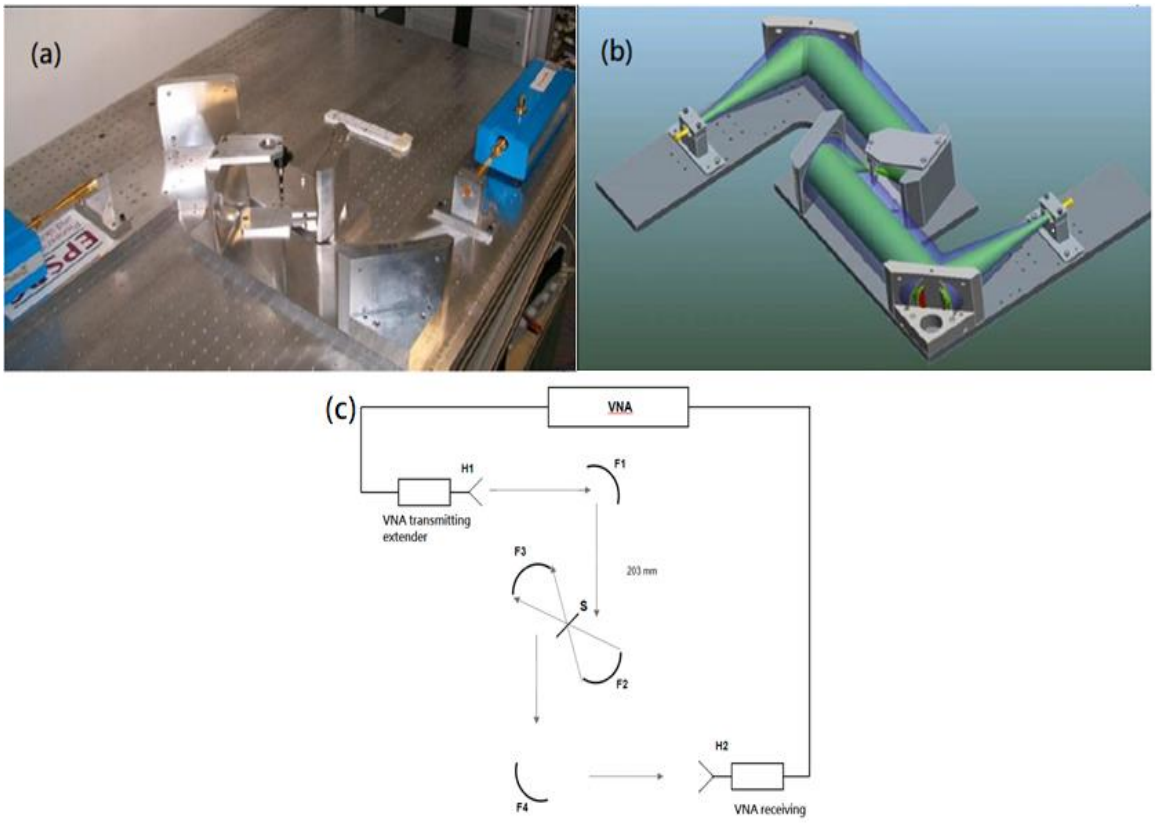

Figure 3.1 (a) A photograph of the physical quasi-optical system in transmission mode [69], (b) A schematic of the 98% Gaussian beam power-level envelope traveling in the circuit [69] and, (c) a schematic of THz QO system from a top view. F1 and F4 are off-axis ellipsoidal mirrors with a focal length of 250 mm. F2 and F3 are parabolic mirrors with a focal length of 100 mm. F1 is used to direct and focus a collimated THz beam onto the parabolic mirror, F2, from the transmitting horn. F2 then focuses the THz beam at the focal point where the sample is located. The emerging signal beam is reflected by another parabolic mirror, F3. The THz beam is refocused by F4 and propagates to the receiving horn H2. The built-in signal generator in the VNA supplies a continue wave (CW) source. Transmit and receive frequency-multiplier modules (blue casings in the photo of 3.1a), provide for CW operation up to 1.1 THz. The QMUL antenna lab currently has four waveguide frequency bands: V (50 to 75 GHz), W (75 to 110 GHz), D (110 to 170 GHz), and H (220 to 325 GHz). Its low-noise performance (0.01 dB) and high dynamic range (> 70 dB) make QO system a versatile tool for measuring the full set of complex-valued S-parameters of given materials.

Sample holder

Several sample holders provide specialized handling for liquid and solid measurement. For liquid samples, a sample holder (A145) purchased from the Bruker company is shown in Fig. 3.2. Two TPX (poly-4-methy1 pentene-1) plates are used as window material due to their extremely low dielectric loss and optical inactivity in the THz spectral domain. The absorption is very low with a refractive index of 1.46 [69]. Both windows are non-

dispersive over the span of 0.1 to 4 THz [69]. The liquid cell consists of two TPX windows, two Teflon gaskets, and a spacer. For introducing the sample into the cell, the gap made by 'spacers' may span from 25 μm to 1 mm.

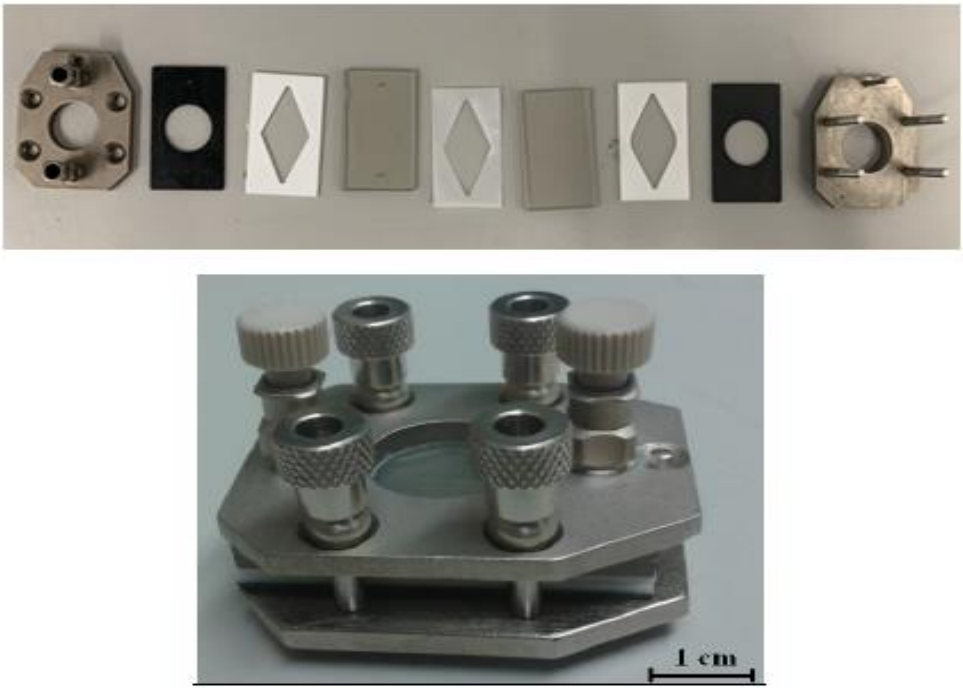

Figure 3.2: The liquid sample holder is purchased from the Bruker company. The liquid cell consists of two TPX windows, two Teflon gaskets (black), and a spacer (white with rhombic aperture). The thickness of spacers ranges from 25 μm to 1 mm. The TPX windows are ultra-low-loss (< 0.1 dB) and optically inactive material, suitable for use as the sample holder in circular dichroism studies.

The 45° wire-grid polariser

The 45° wire-grid polariser is a polariser formed from winding tungsten wires of approximately 5 μm diameter onto a frame with a wire-spacing of 10 μm (see Fig. 3.3 (a)) [70]. The suitable frequency domain is from 50 GHz to 3 THz. Fig.3.3 (b) shows a ray-trace schematic of a free-space signal beam field interacting with a wire-grid polariser. The component of the signal beam with an electric field aligned parallel to the wires, will induce current along the run of the wires. Thus, the polariser acts as a thin metal sheet to reflect that component with some power loss (red arrow in Fig.3.3 (b)). The component of electric

wave perpendicular to the wires does not contribute an electromotive force on free charge carriers and is transmitted (blue arrow in Fig. 3.3 (b)). Since the outgoing signal beam is linearly polarised and perpendicular to the orientation of the wires, the 45° wire-grid polariser placed after sample will only pass a 45° linearly-polarised signal. The cross-polar signal can therefore be calculated from the 45° linearly-polarised signal and the co-polar signal (to be described in 4.2.3).

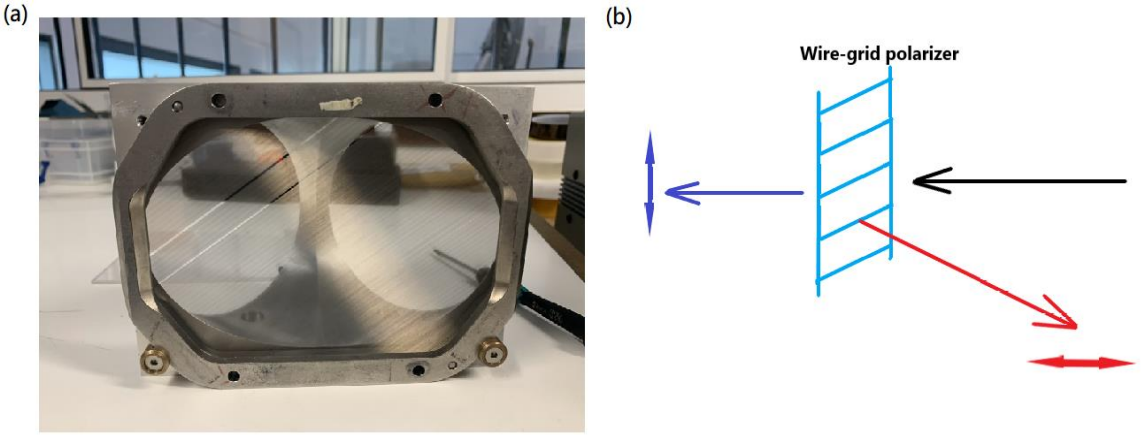

Figure 3.3: (a) 45° wire-grid polariser in the QMUL antenna lab and, (b) a wire-grid polariser reflecting the component of the incident wave (black), parallel to the metallic wire (red), allowing only the component of beam (blue) that is perpendicular to the wires to pass. The frequency domain is from 50 GHz to 3 THz.

Ferrite plate

The ferrite plate is a strontium hexaferrite ceramic material (hexaferrite FB6H supplied by Thomas Keating Ltd.) and is used to rotate the state of linear polarisation of the incident wave through a discrete angle of rotation (due to the Faraday effect described earlier in 2.23). The amount of rotation is determined by the rotation coefficient that is a function of the thickness of the plate. The thickness here is 2.02 mm, and at 300 GHz yields a rotation of 45°. A further description of its performance in this sub-THz QO system is reported by Yang *et al* [66, 71].

3.2 THz Time-Domain spectroscopy

THz TDS is a spectroscopy based on the ultrafast (typically 100 femtoseconds or shorter) pump-probe laser pulse technique [72]. The mathematics of the TDS pump-probe technique is similar to that for interferometry (where two signal pathways, split from a common source, are later reunited (interfered), after each has experienced differing optical events according to the optical paths they took) [73]. TDS depicts the scenario of an ultrafast pump-beam exciting a THz emitter to produce an ultrafast pulse of THz radiation, as the probe-beam gates the detector at sequential timings as the pump beam propagates via a delay-stage. Due to the femto-duration pulse of THz radiation, the average pulse power during the pulse is considerable (approximately 1 W). Since the electric field of THz TDS contains phase information, the biggest advantage is that both amplitude and phase of the THz signal are measured simultaneously, yielding real and imaginary parts of the optical properties of materials [73]. THz TDS has existed in various forms over the past few decades since the Austin switch was developed [74] and has been widely used for the study of materials [75]. Because of the reliability and high sensitivity of ultrafast-laser optical systems, researchers became interested in their biomedical application [76-78]. Numerous types of biological samples such as tissues [79] or small biological molecules [80], were measured using THz-TDS, providing optical properties such as absorption coefficient and refractive index for future study. In the low THz frequency domain (0.1 to 10 THz), THz TDS also serves as an import tool in studying the dielectric relaxational and vibrational modes of liquid, such as water [80].

The schematic diagram of THz TDS in transmission mode is shown in Fig. 3.4.The laser source of the QMUL system is a Class 4 Ti: Sapphire femtosecond, pulsed laser (Mai

Tai, wavelength of operation 750 to 850 nm; repetition rate, 80M Hz; average power, approximately 1 W). The laser source is divided into two beams by a beam-splitter; a 'pump' beam and 'probe' beam. The pump laser illuminates a semiconductor-based (LT-GaAs) photoconductive antenna (PCA). It is comprised of a substrate of 0.5 mm in thickness, with two electrodes separated by a gap of 0.5 mm. A DC bias voltage (200 V) is applied across the parallel-strip electrodes as in Fig 3.5 [81]. When the pump beam illuminates the semiconductor in the gap, free charge carries are induced in the GaAs substrate with the consequent generation of electron-hole pairs. In the photoconductive process, the negative free carriers (electrons, having the higher mobility relative to the positively charged holes (of larger effective mass), are accelerated under the DC bias voltage to produce a transient current with sub-picosecond duration, thereby radiating a THz pulse of electromagnetic radiation [82]. The photocurrent gives rise to the radiated electric field as described by Zhang *et al* [81];

$$\vec{E}(t) = \frac{A}{4\pi r c^2 \varepsilon_0} \frac{d}{dt} I_{pc}(t)\vec{\theta}. \tag{3.1}$$

A is the area in the gap illuminated by the pump laser; ε_0 the vacuum permittivity; r the distance between the field point and the THz source; and, c the speed of light. $I_{pc}(t) = \frac{d}{dt}\vec{p}(t)$ is a transient current proportional to the time-derivative of the dipole moment in an emitter [69].

Off-axis parabolic mirrors (OM) are placed after the PCA to collect and focus THz radiation onto a sample positioned at the focus of OM_2. After propagating through the sample, the THz signal beam is refocused by OM_3. OM_4 focusses onto the nonlinear

electro-optic (EO) crystal (ZnTe) which performs EO sampling (FS-EOS) gated by Δt relative to the probe beam. THz radiation co-propagates with the probe beam in the EO crystal [81, 83]. A delay-stage introduces a time-delay in the pump beam optical path (M_1, M_2, P_1 and P_2) enabling the THz envelope to be sampled in a step-wise fashion [69]. The THz radiation-induced birefringence alters the polarisation of the probe laser in the EO crystal (ZnTe) [84] causing a relative phase delay between the o (ordinary) and e (extraordinary) rays [85] such that the initially linearly-polarised probe laser is transformed to become slightly elliptically polarised. The ensuring rotation of the probe beam is assessed by a quarter-wave plate. In the absence of THz radiation, the probe beam leaving the EO crystal remains linearly polarized, with the quarter wave plate generating circularly polarized light. However, the presence of THz radiation influences the polarisation of light in the EO, transforming the circular polarisation into the elliptical after passing the quarter wave plate. A Wollaston prism then separates the elliptically-polarised light into two orthogonal, linearly-polarised beams. The presence of the THz radiation results in different intensities between the two linearly-polarised beams emerging from the Wollaston prism. These intensities are detected by respective, balanced, photodiodes; i.e. absence of THz will result in a null output.

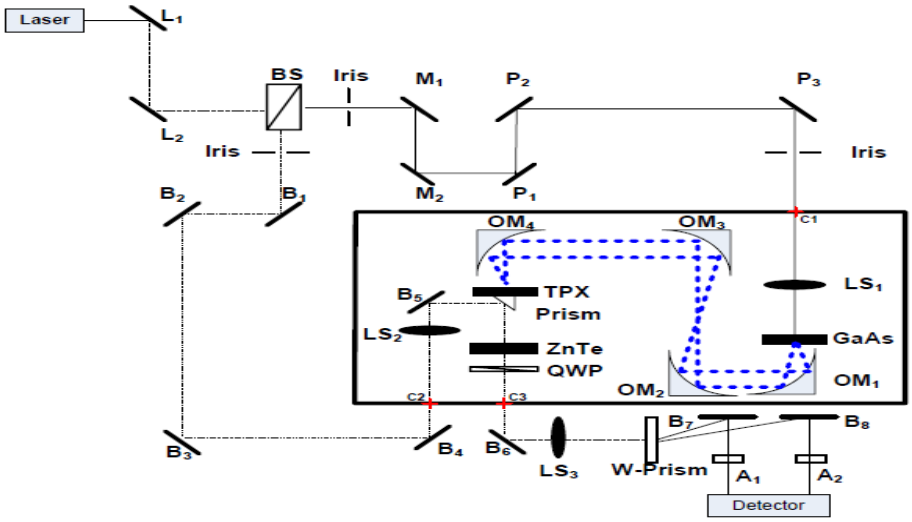

Figure 3.4: Schematic diagram of THz TDS system operating in transmission mode in QMUL. L, M, and P are flat refracting mirrors. BS is the beam splitter; M_1, M_2, P_1, and P_2 are flat reflecting mirrors of the delay stage; LSs are focusing lenses, OM are off-axis parabolic mirrors and QWP is the quarter wave plate. As are attenuators. The blue dash lines show the THz beam path [69]. The pump (solid line) and probe lasers beams (black dash line) are separated by the beam splitter. The pump laser traveling through the translating delay-stage introduces a moving time-delay Δt relative to the probe laser and illuminates the PCA to produce a short time pulse of THz radiation. The sample is placed at the confocus of mirrors OM_2 and OM_3. The THz wave is refocused by OM_3 after passing through the sample. OM_4 then focusses and reflects the wave onto the nonlinear EO crystal (ZnTe) which performs electro-optical sampling gated by Δt relative to the probe laser. THz radiation induces birefringence, transforming the circular polarisation to be elliptical. Elliptical polarisation from the QWP is resolved by the birefringence crystal (Wollaston prism) into two, orthogonal, linearly-polarised, wave components. Balanced photo-detectors are employed to measure the intensity difference between these two component signals.

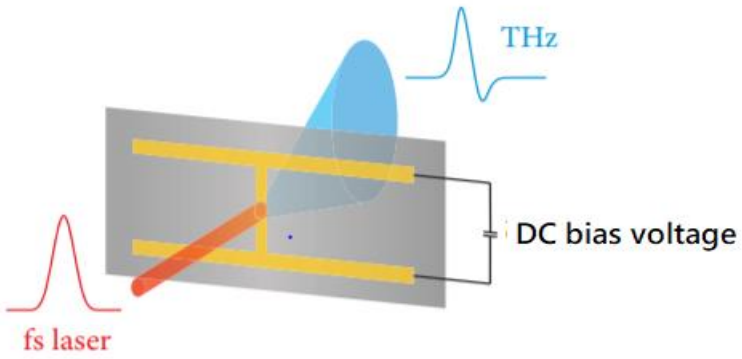

Figure 3.5: The THz pulse generated by photoconductive antenna [82]. A photo-conductive antenna (PCA) consists of two metal electrodes that are fabricated on a semiconductor substrate with a gap between two electrodes. The PCA requires an external DC bias voltage to generate a THz electric field across a small electrode gap. The pump laser illuminates the semiconductor in the gap between the metal electrodes to induce charge carries. The negative charge carriers (electrons) are more mobile and are accelerated under the bias voltage to produce a transient current with sub-picosecond duration resulting in the emission of THz electromagnetic ration.

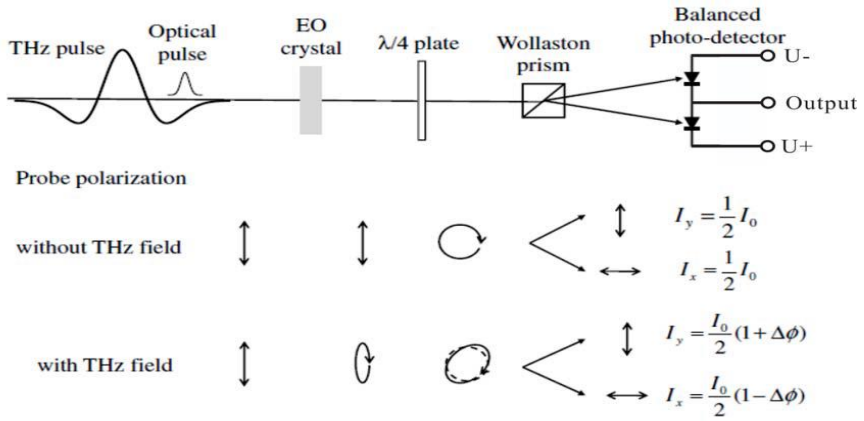

Figure. 3.6: A schematic of free-space electro-optical sampling (EOS) [86]. The EO crystal is non-birefringent in the absence of an THz field. The polarisation of the probe laser shows no change while propagating through the EO. A quarter wave plate is aligned 45° to the linear polarisation and transforms linear polarisation to circular. A Wollaston prism resolves the circularly-polarised laser into two orthogonal, linearly-polarised, beams of equal intensity and consequently no signal is detected in the balanced photo-diode detector pair. In the presence of a THz field birefringence is induced in the EO crystal The co-propagating probe laser becomes elliptically polarised due to the relative phase delay between extraordinary (e) and ordinary (o) beam components. After propagating through the quarter wave plate ellipticity ensues. From the Wollaston prism two, linear, orthogonally-polarised beams have differing intensities. This difference is modulated by the THz beam field and detected by the balanced pair of photo-diode detectors

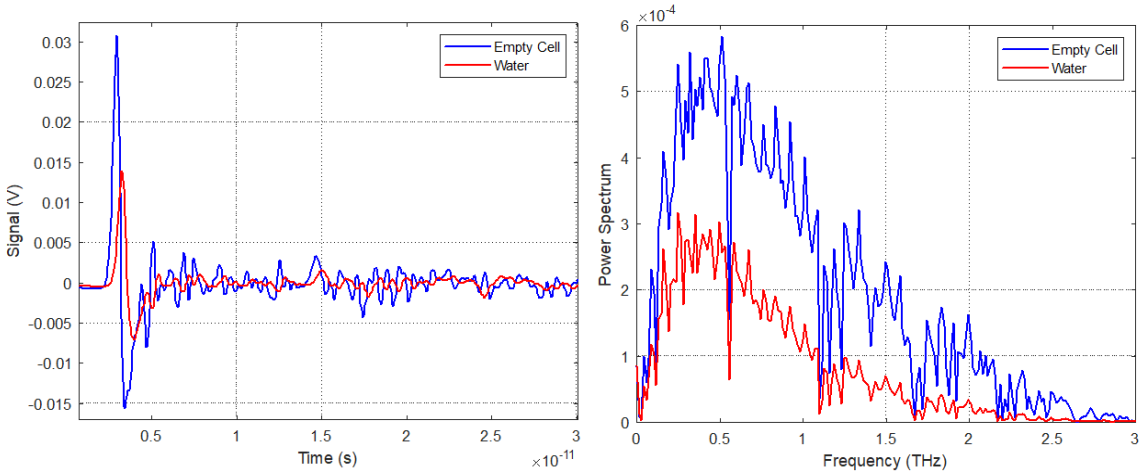

Figure 3.7: The time domain response of empty sample-holding cells and those filled with de-ionised water. (Right) spectra in the frequency domain derived by Fast Fourier Transformation of time-domain data (Left). The relative time delay between water and empty cell is due to the slower velocity of THz radiation in water. The Fourier transformation of the time-domain spectra have sharp dips at around 0.5 THz, a phenomenon attributed to the absorption of water vapor in the air. The reducing amplitude reading of the water spectrum (red) is due to absorption of the probe signal in water. The interaction of the THz electric field with water can be revealed by the ratio of water and reference spectra from a beam path free of water. The scan size of 10.24 mm with 1024 scanning points (time step 10 μm) was set to provide 14 GHz resolution. The biased voltage of the photoconductive antenna was 200 V. The thickness of water was 100 μm.

To conduct liquid sample measurement using THz TDS, an empty sample holder is first measured to provide a reference. The same measuring process is repeated with a liquid

sample injected into the cell. A fast Fourier transform (FFT) maps the measured time-domain signal into the frequency-domain. The time- and frequency-domain spectra of an empty sample cell and one filled with de-ionised water, are shown in Fig. 3.7. The sample and reference spectra are ratioed to reveal the interaction of the THz electric field with the sample. The typical frequency-resolution of the QMUL THz TDS is 14 GHz. The frequency-resolution is related to the number of sampling data points and distance between the data points according to [87, 88]:

$$f_s = \frac{1}{N \times t_s} \tag{3.2}$$

N is the number of data points and t_s is the delay-stage time-step. From Eq. 3.2, a longer scan-time samples more data points, thereby increasing frequency resolution. The scan-length is nominally 10.24 mm with 1024 data points. However, the maximum frequency resolution can be increased to 1 - 2 GHz when required.

Chapter 4

Quasi-Optical Sub-THz Circular Dichroism Spectroscopy of Solvated Myoglobin at D and H Bands

4.1 Overview

Water is well known to cause heavy attenuation of propagating electromagnetic radiation; its specific attenuation at 1 THz is 180 $dB.mm^{-1}$. However, water remains essential in coercing the shape of a biomolecule. Water forms a hydrogen bonding network surrounding the protein, and the coupling between protein and hydration-water by H-bonding triggers strong interactions [89, 90]. Thus, H-bonding plays a critical role in protein folding, stabilization of protein conformation, and protein dynamics. To accommodate the strong absorption by water, the methodology of THz circular dichroism (CD) is explored as a potential workaround to study solvated proteins.

In Chapter 2, the relevant physics of CD was reviewed. Optical CD is widely used to study the secondary conformational forms of protein and these respond differently to different polarizations of light at optical wavelengths. THz CD seeks to sense for expected secondary conformational activity at longer wavelengths. CD is a direct means for studying the dynamics of solvated proteins and it is particularly implemented as described here by means of an analyzing quasi-optical circuit driven by a vector network analyzer (VNA). The VNA supplies the necessary detection sensitivity to compensate for the heavy signal attenuation by water. Implementing THz spectroscopies typically depends on a high-power

laser (e.g. THz time-domain spectroscopy and THz Raman spectroscopy). These THz CD spectroscopies further require signal conditioning schemes to transform from linear–to–circular polarization before irradiating the sample under test. Optical path-length needs to be preserved between sample-versus-reference (background) scans. This typically depends upon a physical enclosure of the THz signal path to maintain a dry (non-absorbing) atmosphere, free of convective turbulence. In this chapter, a simpler apparatus will be introduced. A quasi-optically-based, sub-THz, CD spectrometer is demonstrated on the model protein, myoglobin, over waveguide bands D (110 – 170 GHz) and H (220 – 325 GHz). A significant source of systematic-error is thereby eliminated (and, no less, random-error, that follows when cabling is moved). The 'zero-crossing' points of CD spectra are characterized as they are correlated with traditional far-UV CD spectra for given environmental parameter settings of, for example, pH and sample concentration.

As will be described in section 4.2.2 the benefit of implementing THz CD via QOs lies in the hard and fixed, machined-in accuracy to high-precision tolerances, of optical shaped-surfaces and inter-component path lengths; control of orthogonality of transmit and receive linear polarization states and the kinematic repeatability with which signal-conditioning components, such as wire-grid polarizers and Faraday plates, can be introduced and removed from the circuit to achieve linear–to–circular analysis [71, 91]. To eliminate systematic errors and to receive more phase-stable data, measurements by scheme 1 (involving traditional cross-polar acquisition), and scheme 2 (employing a 45° wire-grid polarizer), will be compared in section 4.2.3

4.2 Experimental Methodology

4.2.1 Sample preparation

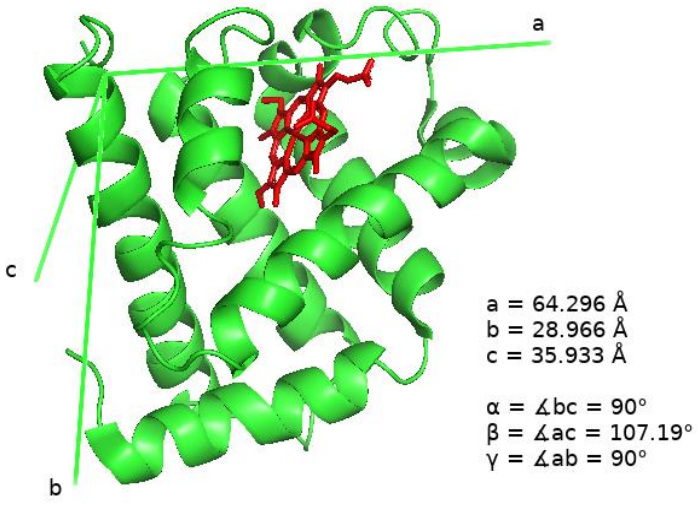

Figure 4.1: The 3D structure of Myoglobin (PDB ID: 1WLA) form equine skeletal muscle, is shown to have monoclinic crystal structure. The green ribbon represents the `backbone` of the molecule and the red represents a heme group. The majority of the backbone structure of Myoglobin is α-helical.

The protein myoglobin (MB) from equine skeletal muscle was chosen as a 'model' system for understanding the results from THz-CD. Lyophilized powder of MB with a purity of 95 to 100% was purchased from Sigma Aldrich and used without any further treatment. 20 mM phosphate buffer (K_2HPO_4 /KH_2PO_4) was chosen as the solvent. By mixing K_2HPO_4 and KH_2PO_4, 1 M of phosphate buffer of pH 8.26 (98ml/2ml), pH 7 (61.5ml/38.5ml), and pH 5.5 (3.5ml/96.5ml) were obtained. All three buffers were verified by a pH meter after mixing and adjusted by adding the appropriate volume of K_2HPO_4 and KH_2PO_4 to achieve the required pH value. Double-distilled water was used to prepare a sequence of buffers from 1 M to 20 mM. The first system studied was 10 mg of MB dissolved in 2 mL of 20 mM phosphate buffer at three pH values (8.26, 7.00, 5.50) without any further purification. To reduce manual error, different pH values of MB solutions were diluted to 1.00 mg/ml and 0.25 mg/ml in separate tubes. MB has been extensively studied

by IR, visible, UV and x-ray light. In the THz spectral domain, the absorption spectra and dielectric properties of MB have been well studied [37, 92]. Fig. 4.1 shows the 3D structure of the equine MB molecule. The green ribbon is the backbone consisting of α-helices to which side chains and other residues are connected. The red structure represents a heme group.

4.2.2 The sub-THz CD QO System

The VNA-driven QO THz CD spectrometer was configured as shown in Fig. 4.2. The VNA has a 10 MHz to 43.5 GHz baseband, and with pairs of transmitting and receiving frequency-multiplier modules, is able to extend frequency-coverage up to 1.1 THz. Transmitting and receiving frequency-multiplier modules were terminated with phase-matched, high-gain, ultra-Gaussian, corrugated horns (H_1 and H_2 in Fig. 4.2). An off-axis ellipsoidal mirror (F1) is set to efficiently receive the beam-field from the transmitting horn and couple it to mirror F2, one of a confocal pair with F3 at whose common focus the sample (S) is situated. F1 and F4, have a 250 mm focal length and F2 and F3 a focal length of 100 mm.

A Bruker liquid-holding cell was used as a sample holder at the con-focus of mirrors F2 and F3. For each filling, 50 μL of the MB liquid was injected into the cell via a pipette. The gap (or the effective thickness of solution), inside the cell is 100 μm. A ferrite plate was introduced prior to the sample, the normal of its broad-face parallel to the beam-axis. The thickness of the plate is 2.02 mm. The function of this leading ferrite plate is to induce Faraday rotation of the vertical, linearly-polarized signal beam from the transmit horn that adds to the cross-polar transmission component (later used, with the co-polar transmission

signal, to synthesise left-handed circular polarisation (LHCP) in particular [47, 71].

In this sub-THz CD QO system, two perpendicular linear transmission signals, $Tcr(\nu)$ and $Tco(\nu)$, were measured and used to synthesis left- and right-handed circularly polarized transmit waves. As noted earlier, the CD spectrum is calculated based on linear co- and cross-polar signals. The equations of right- and left-handed circularly-polarized waves are re-written from chapter 2:

$$T_{+}(\nu) \equiv \frac{1}{\sqrt{2}}\left(Tcr(\nu) - \ iTco(\nu)\right)$$

and

$$T_{-}(\nu) \equiv \frac{1}{\sqrt{2}}\left(Tcr(\nu) + \ iTco(\nu)\right) \tag{4.1}$$

$T_{+}(\nu)$ and $T_{-}(\nu)$ represent the right-hand and left-hand circularly polarized oscillations. The circular dichroism spectrum can then be calculated as

$$\Delta\mathrm{T}(\mathrm{v}) = \ T_{+}(\nu) - T_{-}(\nu) \tag{4.2}$$

4.2.3 Measurement Schemes 1 and 2

The traditional measurement procedure of scheme I measures the cross-polar component of signal propagation in the conventional manner by rotating the receiving horn, H2. The procedure is as follows:

1. **Set up the co-polar measurement as in Fig. 4.2 without the sample and ferrite plat in position S.**
2. **Normalize the free-space background as shown in the reference.**

3. **Introduce the empty Bruker cell and re-set the background to remove interference and etalon reflections originating from the external and internal faces of the empty cell.**
4. **Inject liquid sample into the Bruker cell and position the ferrite plate in front of the sample to measure the co-polar signal, $T_{co}(\nu)$.**
5. **Manually rotate the receiving horn H2 (together with the frequency-extender module), by 90° to measure the cross-polar signal, $T_{cr}(\nu)$.**
6. **Use (4.1) to calculated RHCP and LHCP and obtain the CD spectrum by (4.2).**

Scheme 1 will invariably introduce significant systematic errors. Further, random error is introduced as cables suffer movement away from how they lay during calibration. To avoid such errors, scheme 2 has been devised. It introduces a 45° wire-grid polarizer between F4 and H2 that, with a pair of settings, is used to obtain an equivalent measure of the cross-polar signal component (as is shown Fig. 4.3). A Jones' analysis shows that the linearly-polarized signal at 45° from x-axis, T_{45}, is composed of a mixture of the horizontal and vertical components:

$$T_{45}(\nu) = \frac{1}{\sqrt{2}}\left(Tcr(\nu) + Tco(\nu)\right) \tag{4.3}$$

To determine the cross-polar signal, (4.6) is written as

$$Tcr(\nu) = \sqrt{2}T_{45}(\nu) - Tco(\nu) \tag{4.4}$$

The advantage of this equivalent procedure for determination of cross-polarisation is that cables and receive-horn remain untouched, hence no introduction of phase-noise

and corruption of calibration. Co-and cross-polarisation transmission are used in post-processing analyses to construct LHCP and RHCP transmittances. The procedure of Scheme 2 is:

1. **Repetition of steps 1 to 3 of scheme 1 to acquire the co-polar measurement.**
2. **Remove the sample and ferrite plate from position S and clean the Bruker cell with distilled water.**
3. **Place the empty cell at position S and the 45° wire-grid polarizer in position G1, as in Fig. 4.3, and re-set the background signal.**
4. **Inject liquid sample into the cell and position the ferrite plate in front of the sample to make $T_{45}(\nu)$ measurement.**
5. **Use $T_{co}(\nu)$ and $T_{45}(\nu)$ to calculate $T_{cr}(\nu)$ with (4.7) and obtain CD spectrum by equations (4.3) and (4.4).**

Figs. 4.4 and 4.5 compare the two measurement schemes for a pH 7, 5mg/ml MB liquid sample over D and H waveguide bands. From both figures, the derived cross-polar signals (scheme 2) have been enhanced by the ferrite in order to aid computation of the CD spectrum, with stable values for 'zero-crossings' resulting. Both figures have demonstrated that rotating the receiving horn (and necessarily with it the frequency-extender module), has introduced signal loss and error. At H band, the power loss and difference in absorption between both schemes introduced a frequency shift at each zero-crossing point. Comparison of schemes 1 and 2 demonstrates that the cross-polar signal can be more reliably obtained by $T_{co}(\nu)$ and $T_{45}(\nu)$ yielding a safer CD spectrum.

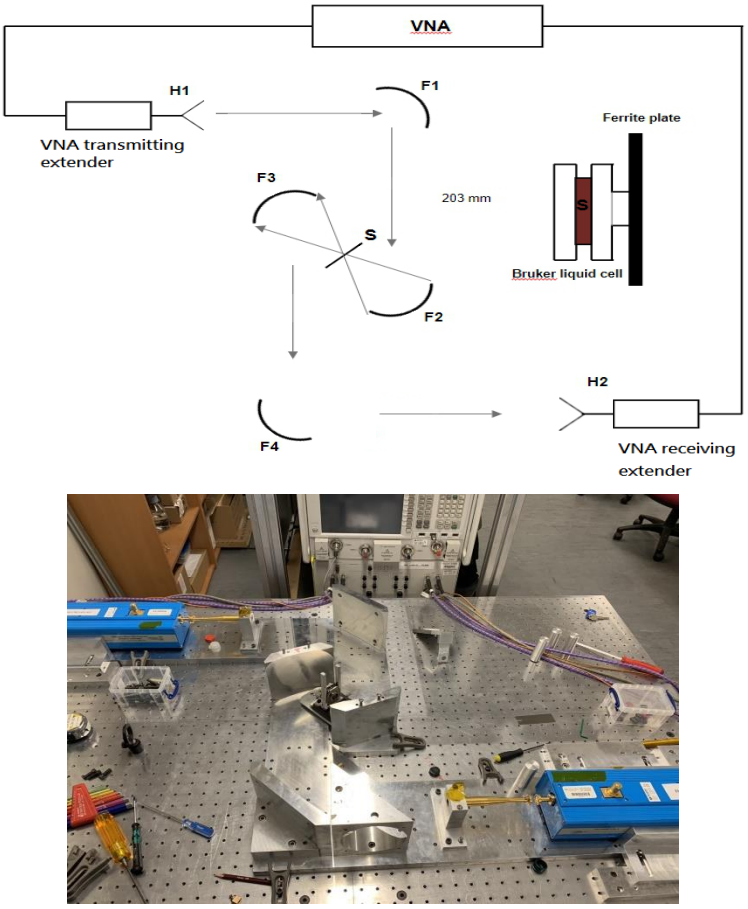

Figure 4.2: Schematic and photograph of the sub-THz quasi-optical apparatus and circuit for implementing circular dichroism spectroscopy. Components F1 and F4 are off-axis ellipsoidal mirrors (common focal length 250 mm; the focal length of the fast parabolic mirrors F2 and F3 is 100 mm). H1 and H2 ultra-Gaussian corrugated horns coupled to frequency extension modules.

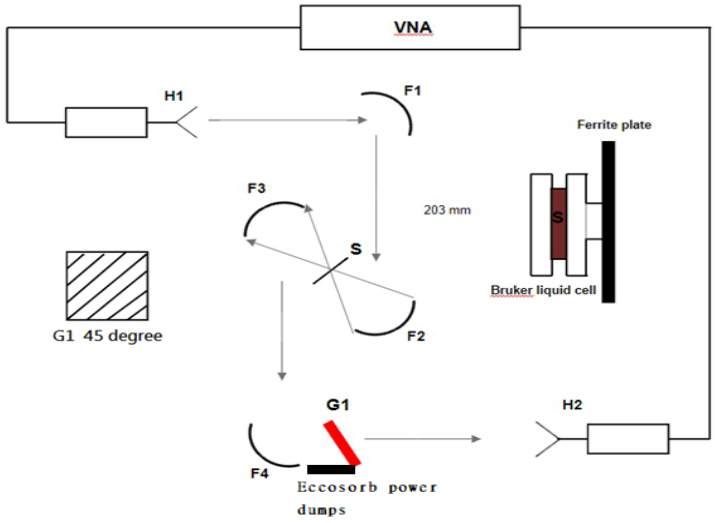

Figure 4.3: Schematic of the Scheme 2 measurement method. A 45° wire-grid polarizer, G1, is introduced between F4 and H2 to measure 45° polarized signal to derive a cross-polar signal that did not entail rotation of the receive horn.

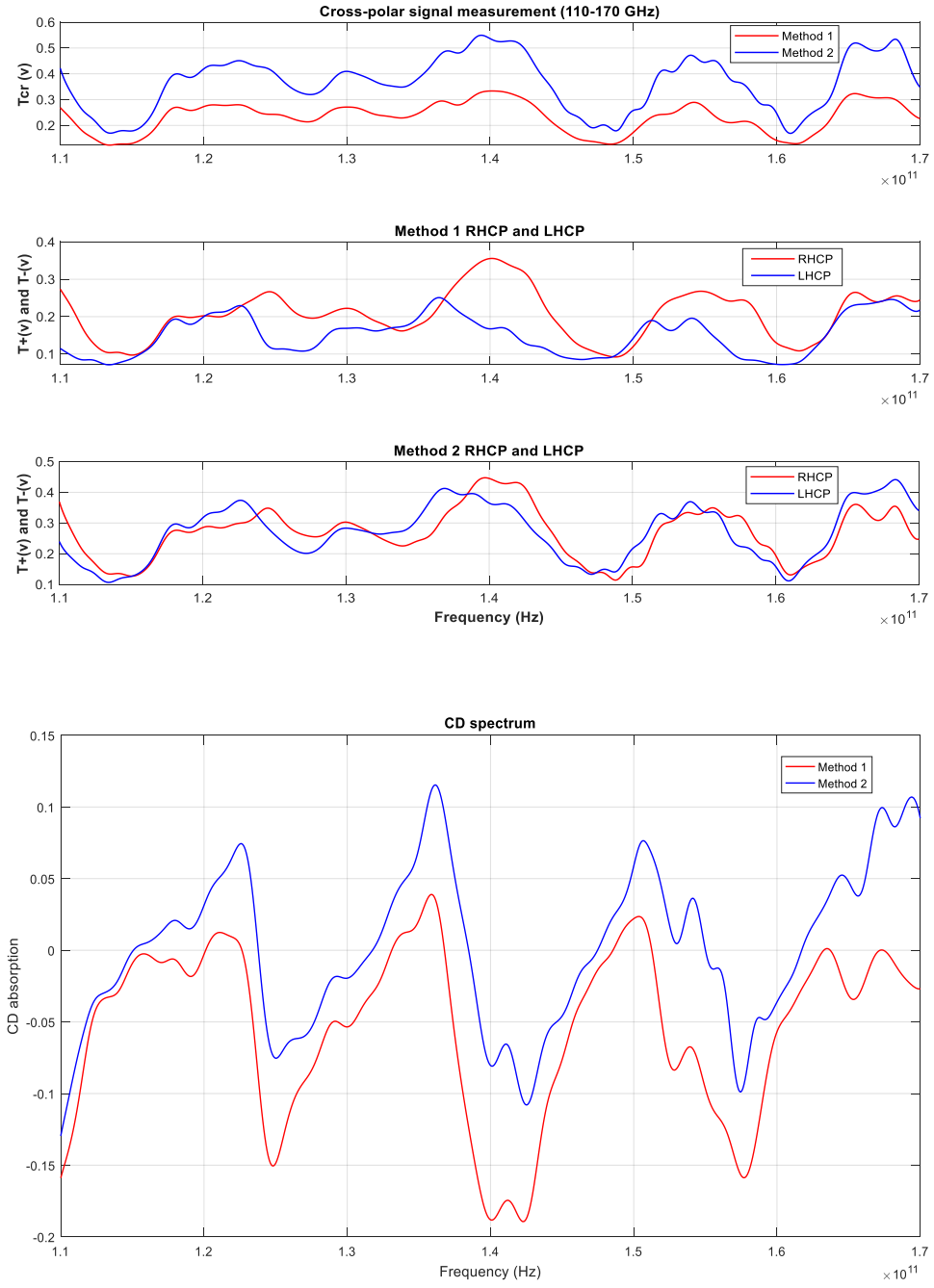

Figure 4.4: D band response of myoglobin (MB) from Schemes 1 and 2. The upper plot shows the traditional and synthesized cross-polar signals. The second and the third plots show the right-hand and left-hand circularly-polarized waves of Schemes 1 and 2. The last plot shows the comparison of CD spectra. By rotating the receiving horn, Schemes 1 has systematic-error to detect cross-polar signal cause diminishing amplitude comparing with Schemes 2.

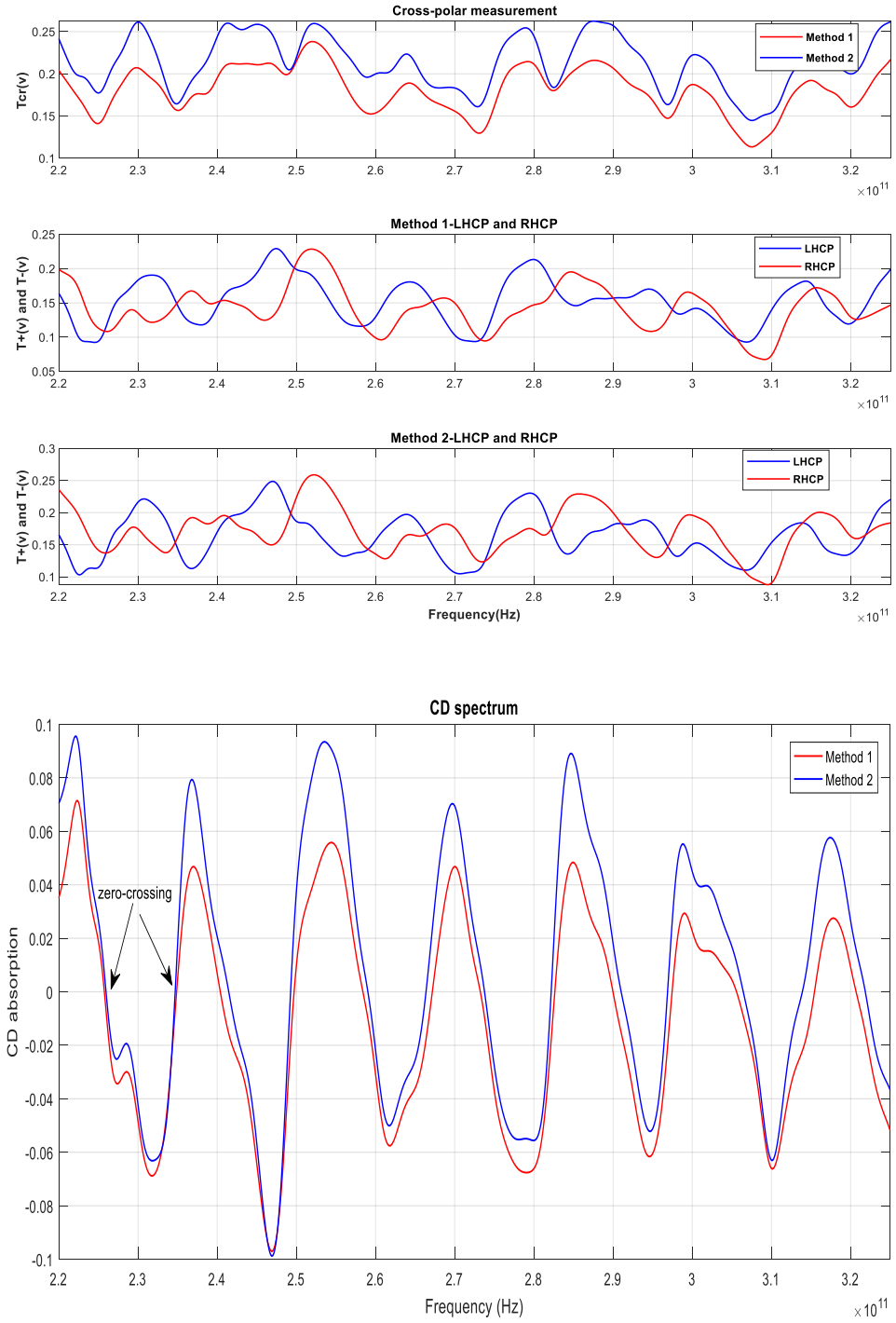

Figure 4.5: H band response from Schemes 1 and 2. The upper plot shows the traditional and synthesized cross-polar signals. The second and the third plots show the right- and left-hand circularly-polarized waves of schemes 1 and 2. The last plot shows comparison of CD spectra. The diminishing amplitude of Schemes 1 cause frequency shift between zero-crossing points.

4.3 Result and Discussion

Far-UV CD spectroscopy is used to independently establish the conformational state of protein secondary structure and thereby interpret sub-THz CD spectra. While far UV CD is sensitive to high-energy activity of side-chain functional groups, sub-THz CD will contrawise be responding to low-energy, group-vibrational modes of the whole backbone of the long-chain molecule. Far-UV CD spectra of MB in 20 mM phosphate buffer at pH 8.26, pH 7.00, and pH 5.50, are shown in Fig. 4.6. The sample thickness is 0.5 mm. The UV CD absorption measurement is based on the Beer's-Lambert-Bouguer law. Concentration level and path-length of cuvette are critical in the UV CD measurement. Normally, for secondary structure study (180-250 nm), 0.10 mg/ml of samples is recommended in a 1 mm path-length of cuvette. The MB solutions were diluted from the 5mg/ml MB samples which were used on the THz CD QO system in the same day. The three spectra show one peak at 192 nm and two troughs at 209 nm and 222 nm. They indicate that the conformation of MB over the three buffer states is α-helix, meaning the protein is stable in those buffers. The diminishing peak at 192 nm and rising features at 209 nm and 222 nm, indicate the influence of pH. When dissolving MB in the buffer below pH 2.50, the strength of H-bonds between the heme group and main chain is weaker and begins to deteriorate. When using buffers of higher or lower pH value relative to neutral, secondary structure will denature into random coils. Although Far-UV CD spectra show clear curves that are directly linked to the protein secondary structure, it is sensitive to sample concentration and significantly influences absorption readings. For MB measurement, concentrations exceeding 0.25 mg/ml reduces S/N for wavelengths longer than 200 nm.

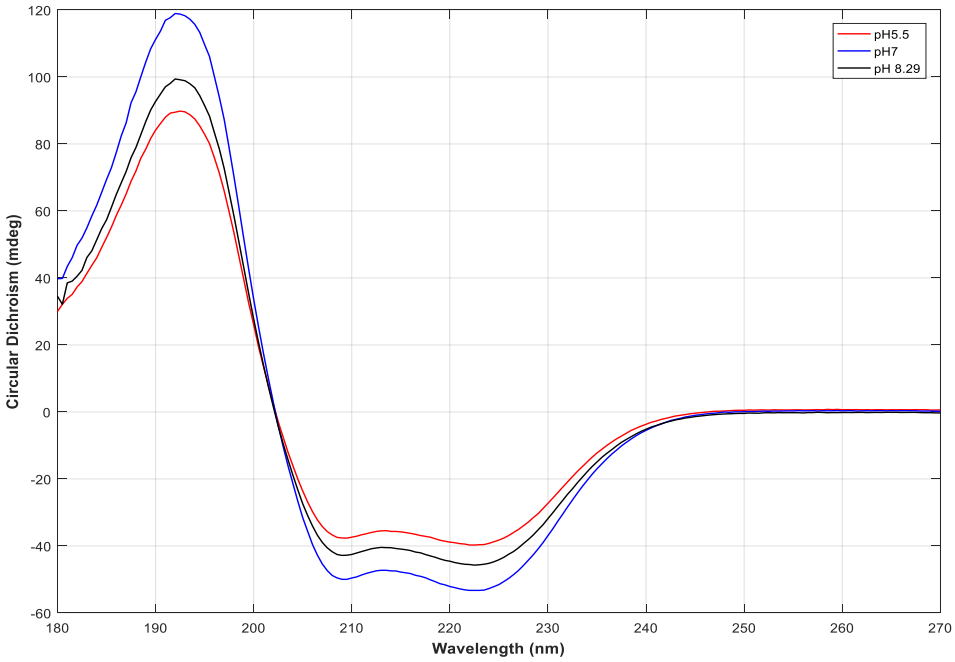

Figure 4.6: Far-UV CD spectra of Myoglobin in 20 mM phosphate buffer for pH 5.50 (red), pH 7.00 (blue), and pH 8.26 (black).

From the comparisons of schemes 1 and 2 above, 2 is to be preferred. All following discussion now concerns scheme 2. The spectral locations of CD zero-crossings are now given to identify samples by acting as their biological 'fingerprint' or 'barcode'. Shifts in zero-crossings result from conformational reconfiguration in the secondary structure of the protein in response to its variations in the state of its environment (e.g. pH, buffer concentration etc). Again, while shorter x-ray and UV wavelengths of light probe the action of light side-chain molecules, longer wavelength THz light probes lower-energy action associated with molecular rotation and coupled-vibrational modes [92]. Such activity of the secondary structure of a protein makes its CD spectrum a unique encoding or register of its condition.

In Fig. 4.7 the red line represents the free-space CD spectrum. Thirteen zero-crossings are noted for the empty cell (whose window material is ultra-low-loss TPX). There is a 100

μm air gap between front and back TPX windows. Different samples induce different levels of absorption of the circularly-polarized wave culminating in different degrees of shift at each zero-crossing point relative to the empty-cell calibration. The THz CD spectrum of the empty cell is dominated by the ferrite plate. Since the sample holder is extremely low-loss and not comprised of an optically-active material, it will not couple to the circularly polarized wave, and so can be considered to establish a free-space CD spectrum. Water, being achiral, is insensitive to the circularly polarized light. Filling the sample-holder with pure water yields an identical THz CD spectrum compared with the empty cell. The reduced amplitude reading for the water CD spectrum follows from water absorption by component co- and cross-polar transmission waves used in construction of the circularly-polarised wave. Significantly, the zero-crossings remain statistically invariant. The thickness of the sample solution is adjusted to be 100 μm. This is a compromise between optimum beam-sample-interaction and S/N. However, by dissolving MB proteins into the buffer (black line in Fig. 4.7), secondary structures of MB show absorption by interacting with the component orthogonal linear polarisations that comprise the THz circularly polarized wave, inducing shifts in where zero-crossings occur. The spectrum of zero-crossings may be regarded as a THz CD bio-signature of MB proteins.

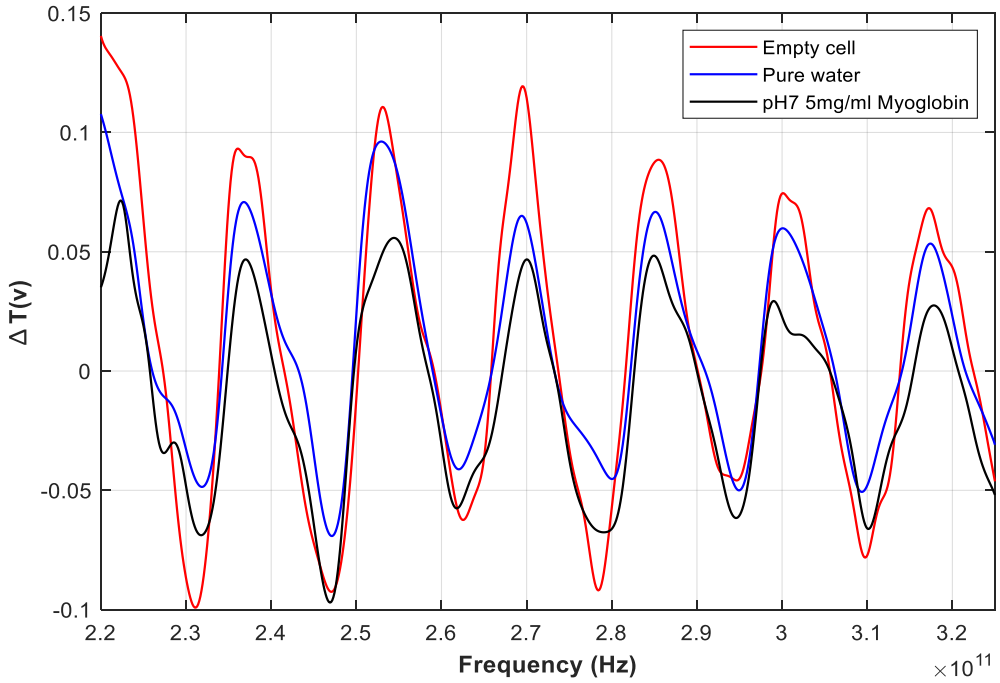

Figure 4.7: CD spectrum of: empty cell (red); pure water (blue); and, pH7.00 5.00 mg/ml the Myoglobin sample (black).

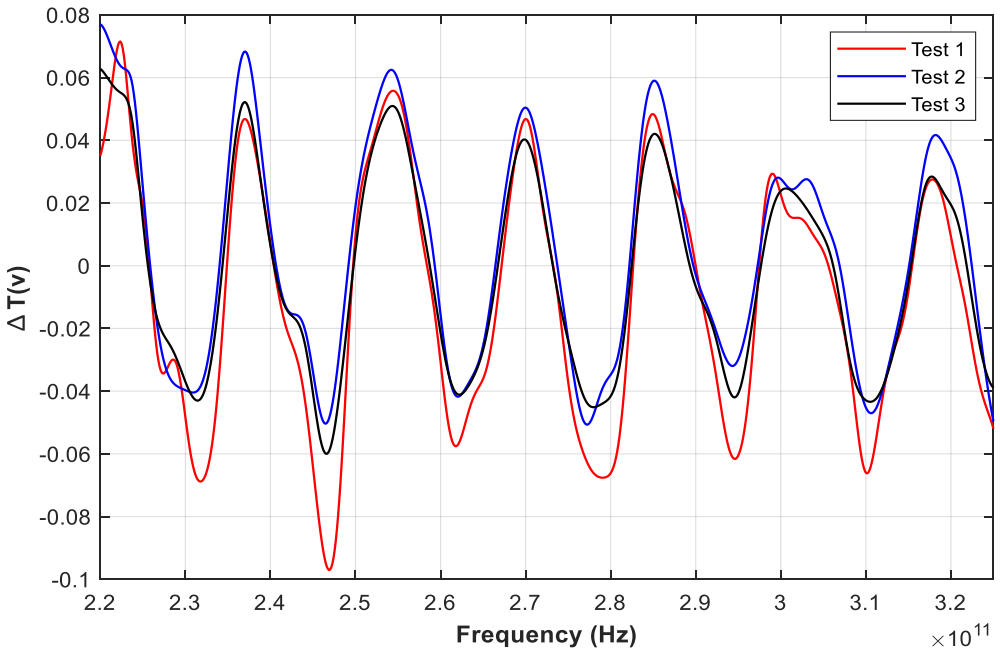

Figure 4.8: Three separate CD measurements of pH 7.00 5mg/mL of Myoglobin for testing the repeatability and errors. In each measurement, the cell was cleaned by pure water and refilled with new sample.

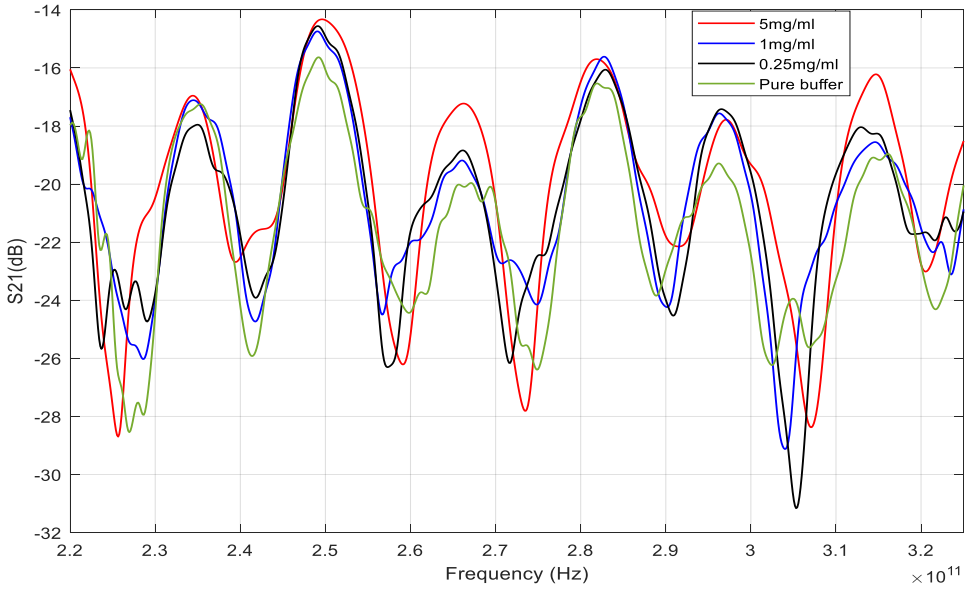

Figure 4.9: S_{21} plots for pH 7 Myoglobin in 5.00 mg/mL (red), 1.00 mg/mL (blue), 0.25 mg/ml (black), and pure pH 7.00 buffer (green) for testing the system sensitivity to low sample concentration.

Before comparing the samples at different pH values of buffer and concentration level, the repeatability and stability of measurements is tested. Fig. 4.8 provides comparison of three CD spectra of pH 7, 5 mg/mL MB solutions. Test 1 corresponds to measurements from a given day. Tests 2 and 3 are subsequent measurements on sequential following days. Test 1, with respect to tests 2 and 3, has an RMS error of 2 and 2.14, respectively. Tests 2 and 3 have RMS error of 0.32. Since the system needed to be re-calibrated, tests 2 and 3 show the smaller error. All the measurements in Figs. 4.10 and 4.11 were done in one day to reduce error. In each measurement, the Brucker cell was washed in doubly-distilled water and refilled with new sample to ensure stability by removing potentially contaminating reside. The measurement procedure followed steps 1 to 4. The sensitivity of the system to samples with a concentration from 0.25 to 5.00 mg/mL needs to be tested. Sushko et al [37], demonstrated that the concentration sensitivity of MB fell from 0.25 to 100.00 mg/ml in the sub-TH spectral domain. Since the ferrite plate is

integral to the measurement process, it is important to test if the system is still sensitive to low sample concentration when introducing the ferrite plate into the transmission measurement path. The S_{21} plot in Fig. 4.9 shows that the system is capable of tracking the difference between each concentration level.

According to far-UV CD spectra, pH values serve as a key parameter in influencing MB secondary structure and its CD absorption reading. The spectra in Fig. 4.6 show the absorption at the first peak drops from 120 to 99 m° (pH 8.26) and 89 (pH 5.50), respectively. The peak at pH 5.5 also shifts from 192.0 nm to 192.5 nm. The drop in absorption at 192 m°, the spiked reading at 209.0 nm and 222.0 nm, and the frequency shift all suggest that the pH value has begun to influence hydrogen bonding inside the folded MB, especially for the buffer with pH 5.50. Figs. 4.9 and 4.10 show the influence of pH values for three different groups of MB liquid samples over D- and H-band mm-wave bands. The pure phosphate buffers suffer the worst phase-noise at D-band. This phenomenon indicates that more obvious frequency shifts between the zero-crossing points in the D band are to be expected. The CD spectra of 5 mg/ml, 1 mg/ml, and 0.25 mg/ml groups show signs of noise above 150.0 GHz which are caused by buffer absorption. By increasing the concentration of MB protein, the influence of buffer absorption is muted and provides seven zero-crossing events in the 5mg/ml group. However, for 1 mg/ml and 0.25 mg/ml, noise becomes significant between 153.0 and 159.0 GHz due to heavy signal attenuation by water.

Each zero-crossing point suffers a frequency shift in the groups of 1 mg/ml and 0.25 mg/ml, their CD spectra approaching that for the pure buffer. In comparison with far-UV CD spectra in Fig. 4.6, pH values of 8.26 and 5.50 slightly reduce the CD reading by

influencing hydrogen bonds inside the folding MB. As noted above, large numbers of water molecules replaced by MB proteins is critical, as such replacement reduces interaction with water molecules and enable more accurate data collection in sub-THz CD measurement. The sub-THz CD spectra for 5mg/ml were analyzed and provided the most stable and clear reading of zero-crossing points which have minimal variation, making this sub-THz CD spectra more reliable when used to track bio-molecular response to its environment. As expected, the comparison of CD spectra in the 5mg/ml group exhibits large frequency shifts occurring at the 2^{nd} (122.9 - 123.8 GHz), 4^{th} (137.9 - 138.5 GHz), 5^{th} (147.5 - 149.1 GHz), 6^{th} (153.1 – 155.0 GHz), and 7^{th} (161.7 - 163.3 GHz) zero-crossing points at D band. The 4^{th} and 6^{th} zero-crossing points for pH 8.26 and pH 5.50 MB samples in both UV and sub-THz CD indicate that the strength of hydrogen bonds inside the proteins' secondary structure have been impacted as aforementioned.

H-band exhibits similar sub-THz CD to D-band (see Fig. 4.11). The sub-THz CD spectra for 1.00 mg/ml and 0.25 mg/ml MB samples are similar to the CD spectra of the pure buffers and have obvious frequency shifts of greater than 1 GHz at each of the zero-crossing points. The spectrum lying between 220 to 230 GHz is strongly affected by the buffer absorption and introduces noise. The influence of buffer absorption is mitigated by raising the level of concentration to 5mg/ml, resulting in 13 zero-crossing points. From the previous description for far-UV CD spectroscopy, pH 5.50 buffer has a stronger influence over the MB secondary structure. At H band, such an influence contributes to obvious frequency shifts at the 2^{nd} (233.8 - 234.9 GHz), 5^{th} (258.2 - 259.5 GHz), 8^{th} (281.7 - 283.1 GHz), 9^{th} (288.9 - 291.8 GHz), 10^{th} (297.4 - 302 GHz), and 11^{th} (305.5 - 307.7GHz) zero-crossing points comparing to that of the pH 7.00 and pH 8.26 MB solutions. However, the

drop of CD reading of pH 5.50 MB solution between 290 to 310 GHz causes shifting at the 9^{th} (290 - 292 GHz), 10^{th} (297 - 302 GHz), and 11^{th} (305.5 - 308.7 GHz) points. It indicates that MB dissolves more readily in pH 7.00 and pH 8.26 phosphate buffers. Although sub-THz CD spectra shows the secondary structures of MB remains α-helix for pH 5.50, pH values of 5.50 and lower (as described in the far-UV CD spectrum), are not native circumstances for MB proteins. By juxtaposing the spectra of pH 7.00 and pH 8.26 MB solution, the frequency shifts at 250, 290, and 297 GHz become more pronounced. The three noticeable frequency shifts indicate that MB secondary structures are also influenced by pH 8.26 buffers, although not as strongly as the solutions for the lower pH values. The positions of all zero-crossing points for 5.00 mg/ml MB in pH 8.26, 7.00, and 5.50 at D and H bands are shown in table 4.1 and 4.2.

The reduced absorption rate at 192 nm and increased values at 209 nm and 222 nm all suggest that the pH value has begun to influence hydrogen bonds inside the folding MB, especially in buffer with pH 5.50. A similar phenomenon is also observed in Figs. 4.10 and Fig. 4.11, showing the pH 5.50 MB sample reporting a more obvious shift than its pH 8.26 MB counterpart. The Far-UV CD spectra from Fig.4.6 have demonstrated that MB is extremely stable when dissolved in these buffers (α-helix**)**, leading to more reliability of sub-THz CD signatures. The Sub-THz CD signatures remain stable even for samples of higher levels of concentration. The highly solvated samples are also measured for biological spectral reading. By measuring proteins and finding their specific sub-THz CD signatures, a database can be constructed to understand the conformational space of proteins and provide insight into the low-energy modes of co-operative vibrations as well as their role in metabolic function.

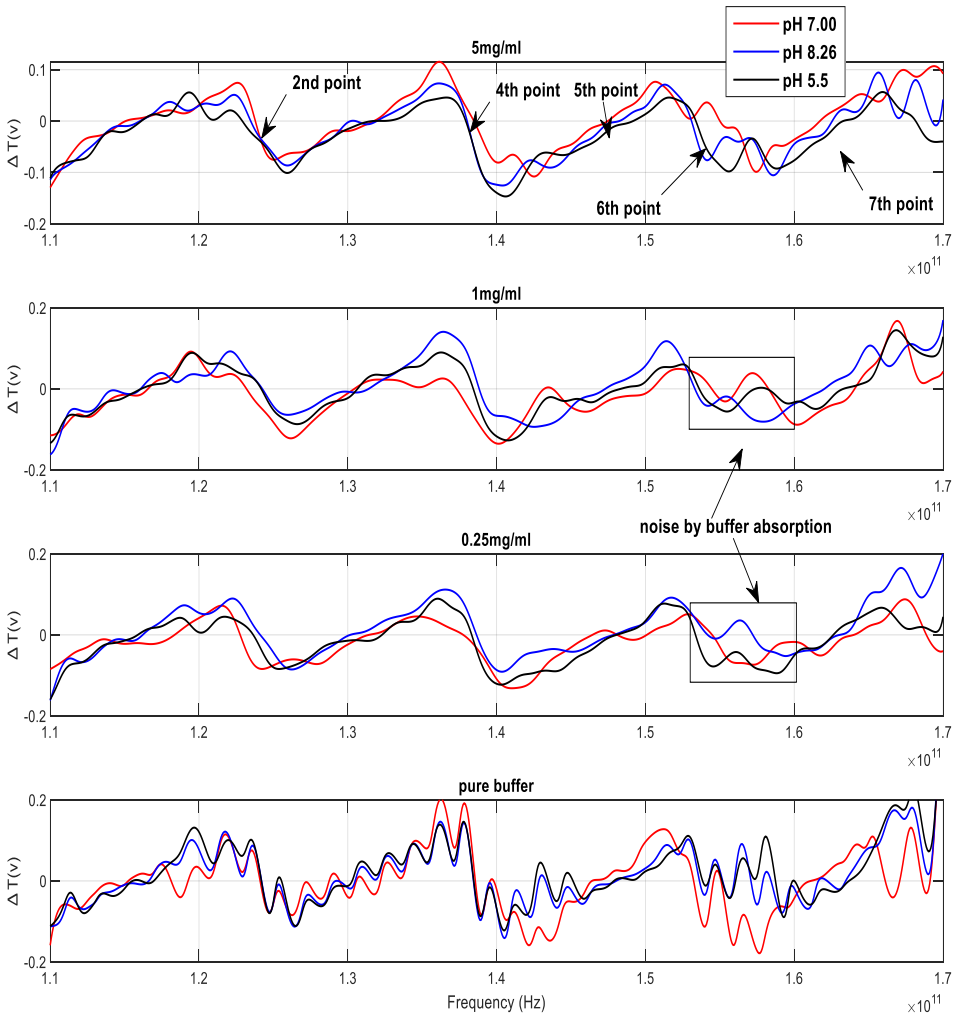

Figure 4.10: Comparisons of sub-THz CD spectra with different pH values at D band (110 – 170 GHz). The concentration level from the first to third plots are 5.00 mg/ml, 1.00 mg/ml, and 0.25 mg/ml, respectively. The last plot shows the comparison with pure phosphate buffer with different pH values. The pure phosphate buffers suffer the worst phase-noise in this band. By dissolving the protein, phase-noise is removed. However, the buffer absorption still strongly influences the CD performance between 150 to 160 GHz.

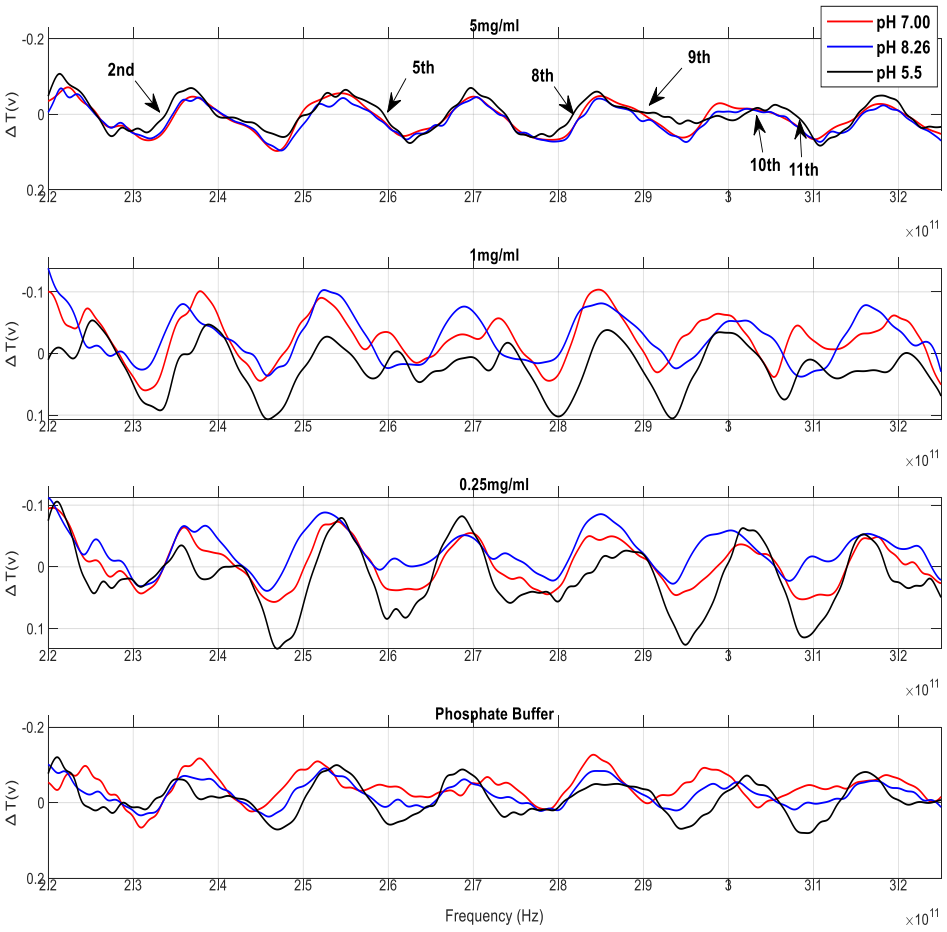

Figure 4.11: Comparisons of sub-THz CD spectra with different pH values in H band (220 – 325 GHz). The concentration level from the first to third plots are 5.00 mg/ml, 1.00 mg/ml, and 0.25mg/ml. The last plot shows the comparisons pure phosphate buffer with different pH values. The different numbers of zero-crossing points indicate that pure buffers still phase noise in the H band. By dissolving the protein, each group of samples show totally 13 zero-crossing points. Since Myoglobin remains folding in three pH values of buffer, the CD spectrum shows same trace in higher concentration level. The frequency shifting at 2^{nd}, 5^{th}, 8^{th}, 9^{th}, 10^{th}, and 11^{th} points caused by the influence of buffer to the protein secondary structure. By reducing the concentration, the CD spectrum of sample have observed frequency shifting between zero-crossing points due to the strong influence of buffer absorption.

Table 4.1: Positions of zero-crossing points for 5.00 mg/ml at D mm-wave band. Frequency values are quoted to five significant figures with an uncertainty of 1% in the least significant figure.

Zero crossing	GHz (at pH 8.26)	GHz (at pH 7.00)	GHz (at pH 5.50)
1st	116.05	115.07	116.10
2nd	123.80	123.46	122.90
3rd	132.00	131.77	132.20
4th	138.02	138.50	137.90
5th	148.25	147.50	149.10
6th	153.10	155.00	153.40
7th	162.55	161.70	163.33

Table 4.2: Positions of zero-crossing points for 5.00 mg/ml at H mm-wave band. Frequency values are quoted to five significant figures with an uncertainty of 1.5% in the least significant figure.

Zero crossing	GHz (at pH 8.26)	GHz (at pH 7.00)	GHz (at pH 5.50)
1st	225.60	225.65	225.70
2nd	234.90	234.65	233.80
3rd	240.50	240.40	239.60
4th	250.87	249.70	250.60
5th	258.20	258.40	259.46
6th	267.30	267.20	266.90
7th	273.00	273.20	273.76
8th	283.10	282.65	281.70
9th	288.90	290.00	291.80
10th	298.17	297.40	302.00
11th	305.50	305.50	308.70
12th	315.70	315.45	315.20
13th	320.50	320.70	320.95

4.4 Summary

A sub-THz Circular Dichroism Quasi-Optical spectrometer operating over the waveguide bands D (110 to 170 GHz) and H (220 to 325 GHz), is introduced. Conventional far-UV CD spectroscopy was used to independently ascertain the secondary structure of Myoglobin. Sub-THz QO CD spectroscopy is aided by the inherent signal and beam conditioning efficiency that is characteristic of quasi-optical circuits and promotes simple sample-handling leading to highly repeatable CD signatures in the study of a proto-typical protein, Myoglobin, with buffers of various concentrations at different pH values. Furthermore, the challenge of working with water-soluble samples was addressed. Though CD is insensitive to water which is achiral, the path-length for the THz probe-beam must be optimized to retain a workable dynamic range to avoid the problem of excessive signal absorption. In contrast, far-UV CD spectroscopy suffers noise when the concentration level of a sample is too high. For far-UV CD, a 5.00 mg/ml Myoglobin solution reports a S/N < 1 for wavelengths longer than 220 nm. However, in the sub-THz CD QO system, the spectra show that higher concentrations aid in securing more stable spectra across both frequency bands tested. So in comparison to far-UV CD spectroscopy, a highly concentrated sample only aids the measurement process. While the initial trials of sub-THz CD QO spectroscopy seem to be suited to liquid biological samples, much supplementary data is required to construct a comprehensive database of solvated bio-molecules.

Chapter 5

Identification of Solvated Proteins by Using Sub-THz Circular Dichroism Spectroscopy from 220 to 325 GHz

5.1 Introduction

Sub-THz CD spectroscopy via VNA-driven quasi-optics has exhibited its stability and sensitivity in the study of conformational change of Myoglobin (MB) in 20 mM phosphate buffers of different pH values. The QO system is capable of detecting sub-THz CD spectra in the concentration range of 0.25 – 5.00 mg/ml with the sample volume having a thickness of 100 μm. As mentioned in Chapter 2, the concentration of protein in the solution and the sample thickness in far-UV CD absorption measurement are based on Beer's-Lambert-Bouguer law (Eq. 2.23). However, the sub-THz CD absorption measurement is a different story. Unlike far-UV CD measurement, sub-THz radiation interacts with low-energy molecular motion (such as rotation and torsion) [93] and protein vibrational modes [26, 68, 76]. To study solvated proteins using sub-THz CD spectroscopy, the CD absorption of protein must be greater than the water attenuation so the difference of CD signals can be detected. Chapter 4 established that samples of high concentration are more suited in identifying MB conformational changes in buffers at different pH values. The pathlength of the liquid sample cell is as important as the protein concentration level. If the liquid sample is too thin, the optical path length is diminished for sensing CD at mm – submm wavelengths; the interaction length being sub-optimal.

Buffer selection too is critical. Recall from Chapter 2 that some buffer systems had strong absorption in far-UV CD measurements, unless diluted to extremely low molar concentration, a phenomenon that is best demonstrated by the phosphate buffer. Fig. 5.1 shows the CD spectra of an empty cell, 20 mM pH 7.00 phosphate buffer, and pure water. As mentioned in chapter 4, the THz CD spectrum of the empty cell is determined by the ferrite plate. Since the sample holder is extremely low-loss, and not comprised of any optically-active material, it will not couple to a circularly polarized wave. Thirteen zero-crossings occur in the 230 – 325 GHz interval. With the cuvette filled with water alone, the 13 zero-crossings are unaltered, as expected (water being achiral). However, pH 7.00 phosphate buffer registers a CD response. The salt dissolved in the buffer increase absorbance and affects THz CD spectrum [55].

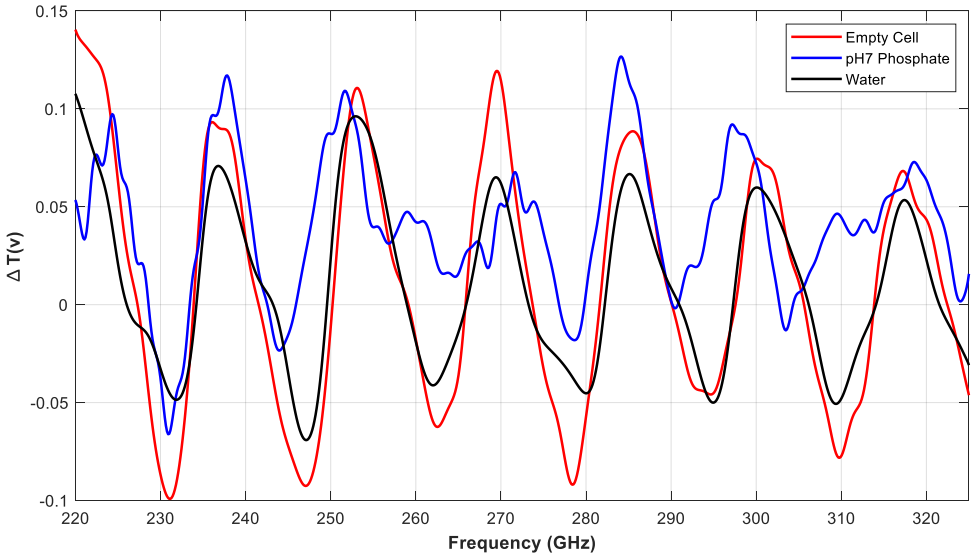

Figure 5.1: Sub-THz CD spectra of 20 mM pH 7.00 phosphate buffer and pure water. The salt in the buffer can affect THz CD spectrum by the absorbance increase in higher frequencies [5]. THz CD spectrum of Phosphate buffer provides only 8 zero-crossing points due to the strong absorption. In Far-UV CD study, phosphate shows low absorbance in shorter wavelength [55]. In sub-THz range, the spectrum indicates that phosphate buffer has strong absorbance and provide only 8 zero-crossing points in the H band.

The absorption of solvated proteins has been well explored in the sub-THz and THz spectral domains [37, 94-97]. Although solvated protein reports higher absorption

efficiency than water from 2 to 3 THz [96], strong water absorption remains a huge obstacle for conventional THz spectroscopic methodologies. But according to Fig. 5.1, to avoid interference of buffer in studying sub-THz CD signatures of solvated proteins, water seems to be a better buffer option. The QO system is a useful technique in probing solvated protein from 0.25 to 5.00 mg/ml [94]. By adopting sub-THz CD QO spectroscopy in solvated protein study, their THz CD spectra can be taken as 'fingerprints', their differing CD absorptions giving uniquely varying frequency zero-crossings. Increasing the sample thickness, the number of zero-crossing points remains constant. Sub-THz CD therefore suggests a possible route through in physically interpreting conformational variation in response to variation in environmental changes experienced by the dissolved protein.

5.2 Measurement Method

5.2.1 Protein

Myoglobin (MB), bovine serum albumin (BSA), and Papain were chosen as model systems for demonstrating the effect of THz-CD. All proteins were purchased from Sigma Aldrich in the form of lyophilized powder and used without further treatment. MB powder from the equine skeletal muscle is 95 to 100% pure. The purity of BSA is from 98 to 100 %. Papain is extracted from papaya juice. The entire secondary structure of MB consists of 8 α-helix segments of various sizes (approximately 75 %) with no β-form and random coil, making it an ideal standard protein for sub-THz CD analysis. Secondary structure of BSA consists of approximately 54% α-helices and 18% β-form [95]. Secondary structure of papain consists of approximately 25% α-helices and 21% of β-sheets [96]. All proteins are soluble in water.

Doubly-distilled water was used as the buffer. Samples were separated into two groups based on their concentration levels, namely 10 and 5 mg/ml. All 10 mg measures of the proteins were first dissolved in 1 ml of water without any purification. Samples used in the measurement were prepared at two levels of concentration; 10mg/ml and 5mg/ml. To reduce manual errors, all 10mg/ml protein solutions were diluted to 5mg/ml in separate tubes. At THz frequencies, the absorption rate, dielectric response and solvent dynamics of MB [37, 78, 97], and BSA [37, 68, 97, 98] have been well studied by various experiment methods. Moreover, THz EM absorption of MB, and BSA solution at different levels of concentration was extensively studied by Sushko *et al* using a QO system that measured over the frequency span 220 to 325 GHz [37]. Papain has been less frequently studied by the THz community. Papain has instead been the subject of study regarding its molecular structure in the IR and UV bands of the EM spectrum [99, 100].

5.2.2 Measurement method

A Bruker liquid cell was used as the sample holder. Three PTFE (Polytetrafluoroethylene) spacers of different thickness (100 µm, 310 µm, and 500 µm), were placed between two TPX windows to create a gap (establishing the effective thickness of the solution sample). For each filling, the sample volumes were 50 µL, 150 µL, and 250 µL, respectively. Between each measurement, the cell was washed with distilled water and the system background signal re-taken.

The measurement method by VNA-driven sub-THz CD QO spectroscopy followed the aforementioned, schemes 2, in the previous chapter. By introducing the ferrite plate before the sample, the co-polar linear polarisation component (i.e. ‘vertical’ component with respect to the plane of the optical table), T_{co}, was measured. For cross-polar (i.e.

horizontal), measurement, the empty cell and the 45° wire-grid polarizer are first respectively placed in positions S and G1, with the background signal reset as described earlier in relation to Fig. 4.3. With injection of the liquid sample into the cuvette, the 45° signal was then recorded with theVNA. A Jones' analysis showed that the linearly-polarized signal at 45° from x-axis, T_{45}, was composed of a mixture of horizontal and vertical components [73]. The relation between T_{cr} and T_{45} is expressed as:

$$Tcr(\nu) = \sqrt{2}T_{45}(\nu) - Tco(\nu) \tag{5.1}$$

The expressions for right- and left- handed circularly-polarized waves are, respectively:

$$T_{+}(\nu) \equiv \frac{1}{\sqrt{2}}\left(Tcr(\nu) - iTco(\nu)\right)$$

and

$$T_{-}(\nu) \equiv \frac{1}{\sqrt{2}}\left(Tcr(\nu) + iTco(\nu)\right) \tag{5.2}$$

The CD spectrum is then computed as

$$\Delta\mathrm{T}(\nu) = T_{+}(\nu) - T_{-}(\nu) \tag{5.3}$$

5.3 Result and Discussion

To minimise systematic error, the experimental methodology reported in Ch 4 was again followed so that signal cables need not be disturbed. RMS error between each test was consequently kept below 1. (Recall from Chapter 4, Fig. 4.8 that, applying a re-calibration which involved a change of conditions between the background and sample

measurement because signal-cabling was moved, can put RMS > 2). To obtain reliable and accurate sub-THz CD signatures, each sample was repeat-measured three times as in Fig. 5.2. For MB, the result of test 3 have observed zero-crossing shifts at 1st (0.26 %), 5th (0.25%), 7th (0.32%), 9th (0.75%) 11th (0.36%) and 13th (0.81%) points comparing with test 1 and test 2. For BSA, observed frequency shift has noticed at 9th (0.57% between test 1and 2 and 0.77% between test 1 and 3) and 13th ((0.39% between test 1and 2 and 0.58% between test 1 and 3) points. A resultant protein CD signature is given as the average of the three measurement spectra.

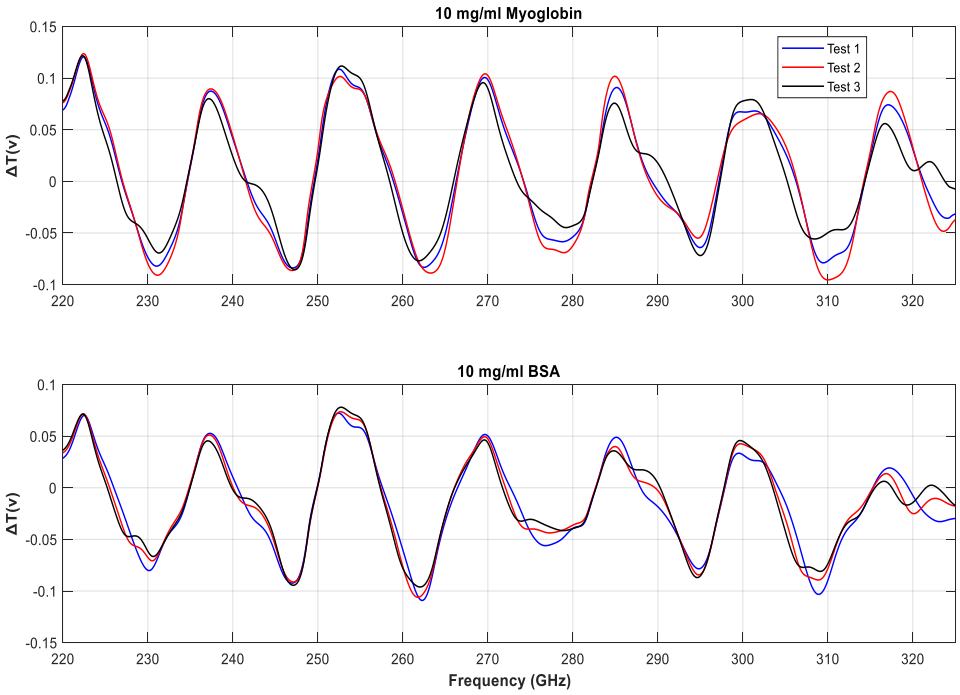

Figure 5.2: Sub-THz CD spectra of myoglobin and bovine serum albumin in three individual tests. Tests 1, 2, and 3 are separate repeats in the same day to system stability. Between each measurement the liquid sample-holding cell is washed with distilled water with the background signal reset. The average of the three spectra is calculated for a more reliable and accurate sub-THz CD spectrum. The final protein CD signatures are then used in later comparison.

As mentioned in section 5.1, the concentration and optical pathlength of the sample are critical to the sub-THz CD measurement. Higher concentration provides adquate CD

obsorption. However, if the sample is too thin, the proteins will not provide enough CD absorption and the sub-THz CD spectra will be indistinguishable. However, if the sample is too thick, the CD signal will be too weak for any meaningful interpretation. With the appropriate pathlength, the sub-THz CD spectrum of a sample of lower concentrated will be much closer to that of water. Fig. 5.3 and 5.4 show the change in CD spectra of 5 and 10 mg/ml MB and BSA soiutions with the sample thicknesses of 100, 310, and 500 µm.

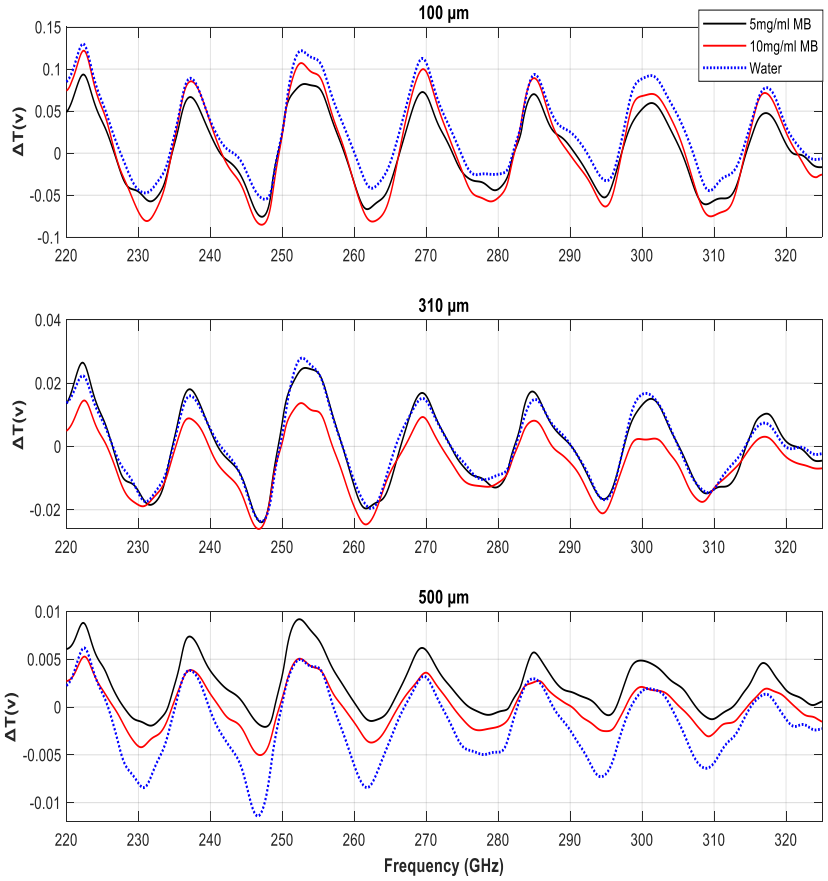

Figure 5.3: Sub-THz CD spectra of 5 and 10 mg/ml myoglobin solutions with sample thicknesses of 100 (upper), 310 (middle), and 500 µm (bottom). In the upper plot, myoglobin solutions of various concentration levels show no observable frequency shift at zero-crossing points due to the shorter pathlength. This phenomenon indicates that 5 mg/ml and 10 mg/ml myoglobin solutions have the same CD protein absorption to the CD signal while traveling inside the sample. By increasing the sample thickness to 310 µm, the protein concentration must be raised to 10 mg/ml to provide enough CD absorption. Since 5 mg/ml Myoglobin solution does not provide enough CD absorption for the proteins, the CD spectra of protein (black) solution shows no observable frequency shift with respect to that of the water (blue).

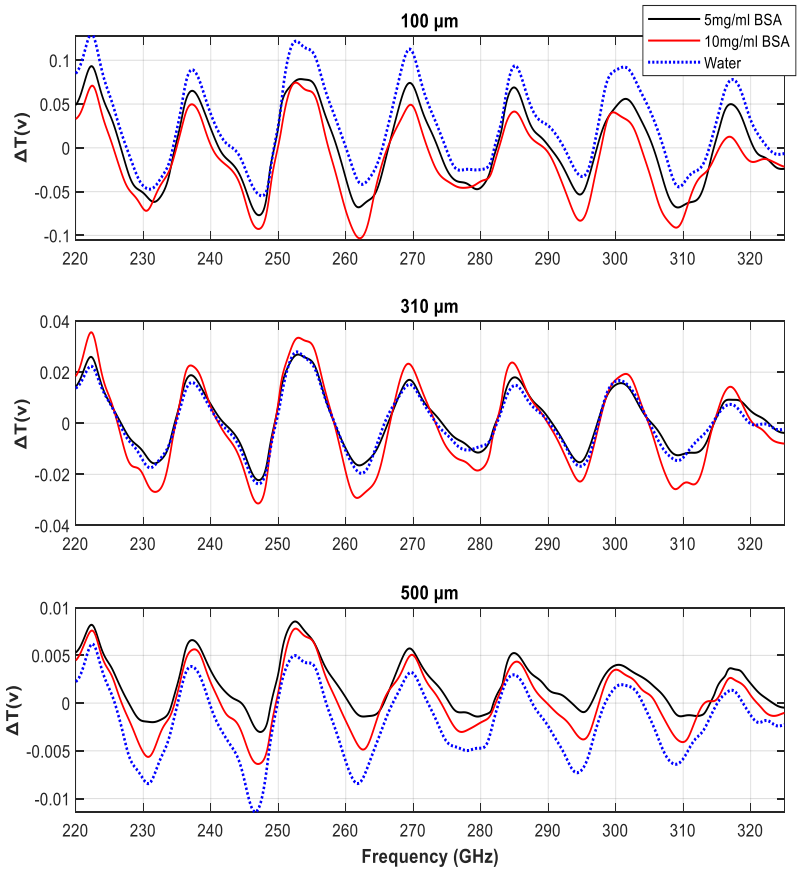

Figure 5.4: Sub-THz CD spectra of 5 and 10 mg/ml bovine serum albumin solutions with sample thicknesses of 100 (upper), 310 (middle), and 500 µm (bottom). In the group of 310 µm, the results of bovine serum albumin solutions are compared with that of the myoglobin solutions in Fig. 5.3. The CD spectra of low concentrated sample show no observed frequency sift comparing with that of the pure water due to less absorption of the protein's CD signal. In the group of thin sample thickness, unlike myoglobin solutions, the frequency shifts at each zero-crossing points are observed, indicating that the concentration level of bovine serum albumin can strongly influence the CD spectra. By comparing with the myoglobin as in Fig. 5.6, the spectrum of 5 mg/ml bovine serum albumin and myoglobin solutions reports no distinguishable features between different secondary structures. Such finding indicates that higher concentration of bovine serum albumin solutions are required to provide accurate CD spectra reading of its secondary structure.

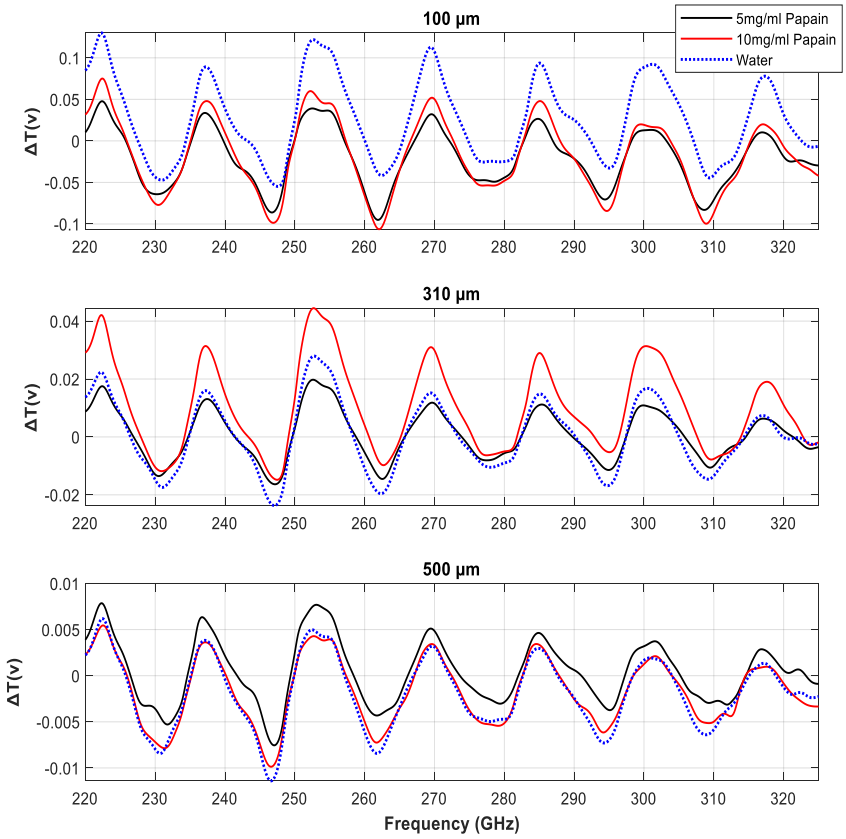

Figure 5.5: Sub-THz CD spectra of 5 and 10 mg/ml Papain solutions with sample thicknesses of 100 (upper), 310 (middle), and 500 µm (bottom). In the upper plot, the CD spectrum of Papain solutions have no observed frequency shift due to the same protein CD absorption. By increasing the sample thickness by three times, the 5 mg/ml Papain solution has no observed frequency shift comparing with that of the pure water due to the weak CD absorption provided by the low concentration of proteins. In the bottom plot, the CD spectrum of water and 10 mg/ml Papain solution indicate that the concentration level needs to be increased for thicker samples to avoid strong water attenuation.

In the group of 100 µm thickness, spectra of 5mg/ml and 10 mg/ml MB, and Papain have no obvious shift at each frequency of zero-crossing. Solutions of 5 and 10 mg/ml BSA, however, have obvious frequency shifts. Comparing spectra with water, shifting at some zero-crossing points can be observed. When the pathlength increases to 310 µm, spectra of 5 mg/ml MB, BSA, and Papain solutions barely shift when compared with water. It shows that 5 mg/ml of samples is too dilute to provide a viable CD measurement. By doubling the concentration, the number of proteins in the optical interval is enhanced sufficently to

yeild a CD measurment. The concentration of 10 mg/ml is more appropriate with samples occupying a 310 μm gap. With respect to water, 10 mg/ml of Papain solution has no shift due to the water attenuation. The secondary structure of Papain contains only 25% α-helices and 21% of β-sheets. In comparison with MB and BSA for the500 μm gap, Papain is measured to be less sub-THz CD active due to the lower percentage of α-helices. For BSA (Fig. 5.4 third plot), over 50 % of α-helices and 18 % of β-sheets can still provide enough sub-THz CD absorption under strong water attenuation. Fig. 5.3, 5.4, and 5.5 show that 10 mg/ml protein solutions are more sub-THz CD active. Although spectra of 5 and 10 mg/ml protein samples display no obvious spectral shifts in zero-crossings for the 100 μm gap, there is nonetheless discernable CD activity with respect to the baseline water measurement. In the group for the 310 μm gap, only 10 mg/ml protein solutions will be considered because the sub-THz CD spectra of lower concentrations show no obvious activity. In the group for the 500 μm gap, Papain shows no CD activity for 10 mg/ml, while MB and BSA, do. In this case, both 5 and 10 mg/ml of samples will be considered for further study.

Fig. 5.6 shows the comparison of sub-THz CD spectra of MB, BSA, and Papain for a sample gap-width of 100 μm. 5 mg/ml of MB and BSA have similar sub-THz CD spectra and no discernable frequency-shift at each zero-crossing point. Recall that the secondary structure of MB consists of 75 % α-helices and no β-sheets while BSA consists of 54 % α-helices and 18 % of β-form. At this concentration, no CD activity is registered. Papain contains less than 50 % α-helical segments and β-sheets. Over 50 % of enzyme structures in secondary conformation result in strong absorption, especially above 290 GHz. In Fig. 5.5, spectra of 5 and 10 mg/ml Papain show no obvious frequency-shift in zero-crossings.

By increasing concentration, response of BSA moves to lower frequency to overlap with that of Papain. However, shift in zero-crossing frequencies can be detected at 289 and 319 GHz.

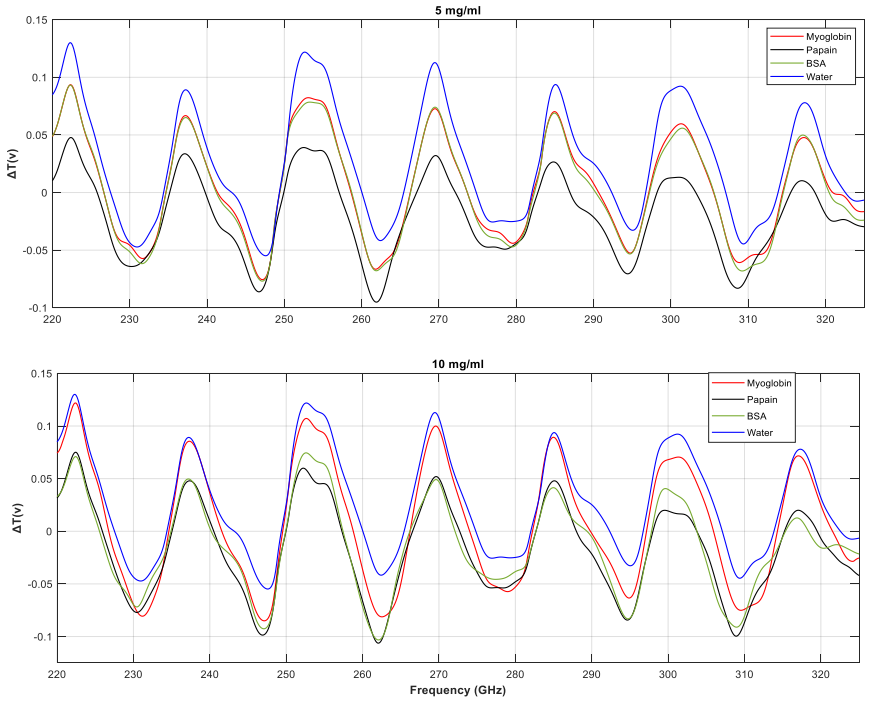

Figure 5.6: The upper plot shows the sub-THz CD spectra of 5 mg/ml of myoglobin, bovine serum albumin and papain with a pathlength of 100 μm; the bottom shows comparison of sub-THz CD spectra for a doubling of. In the group of low concentration, myoglobin and bovine serum albumin solutions provide the same CD spectrum as they have the same absorption of protein's CD. This phenomenon indicates that CD signals of two samples are not distinctive enough to provide frequency shifting at each zero-crossing points. By doubling the concentration level, the myoglobin solution containing higher ratio of α-helices has a observed frequency shift at least 0.5 % at each zero-crossing pints comparing with that bovine serum albumin and papain. For bovine serum albumin and papain solutions, since bovine serum albumin contains almost twice as many α-helices as Papain, large frequency-shifts at the 9^{th} (1.38 GHz) and 13^{th} (1.15 GHz) zero-crossings have been observed.

In Fig. 5.7, the upper plot shows that sub-THz CD spectra for all proteins overlap and have no discernable optical activity. The second plot in Fig. 5.7 shows that by increasing the concentration level, spectra of protein solutions show a frequency shift at most of the zero-crossing points. Comparing BSA to its water-based background, frequency shifts at 266 and 320 GHz are distinct. The amplitudes of both spectra also indicate the existence of solvated BSA. The increasing amplitude of BSA shows that 10

mg/ml of protein molecules mitigate the attenuation by water. The MB solution shows strong CD due to the α-helical segment in its secondary structure. Comparing the three proteins, a lower percentage of α-helixes are expected to show less sub-THz CD.

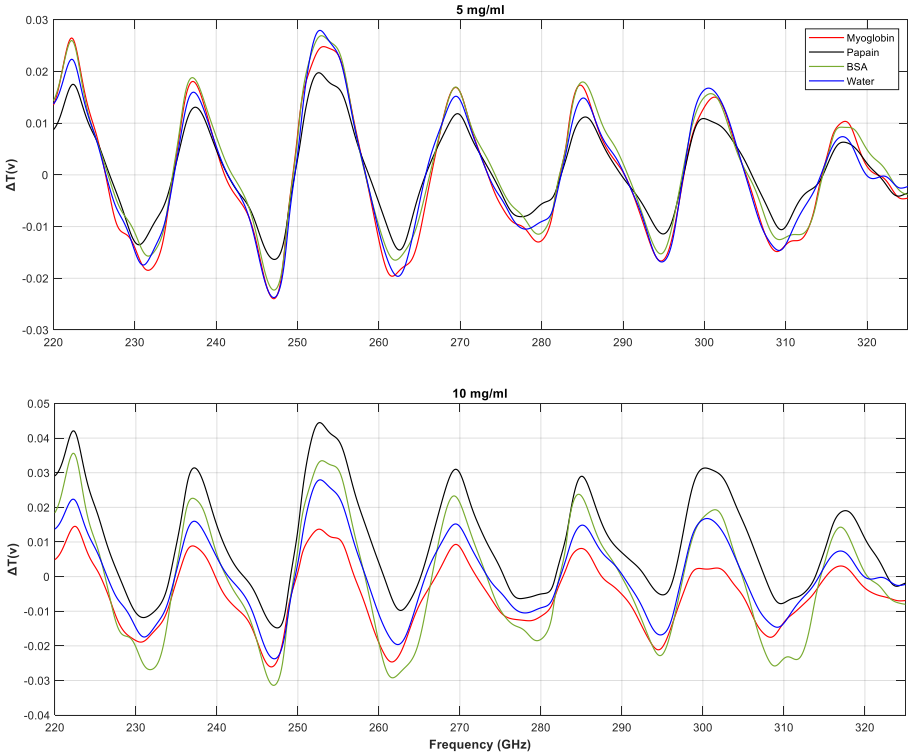

Figure 5.7: Sub-THz CD spectra of 5 mg/ml (upper) and 10 mg/ml (bottom) myoglobin, Papain and bovine serum albumin solutions with pathlength of 310 μm. In the group of low concentration levels, the CD spectrum of three samples show no observed frequency shift due to the weak absorption of protein's CD. By raising the concentration level, three CD spectrum report observable frequency shifts at each zero-crossing point.

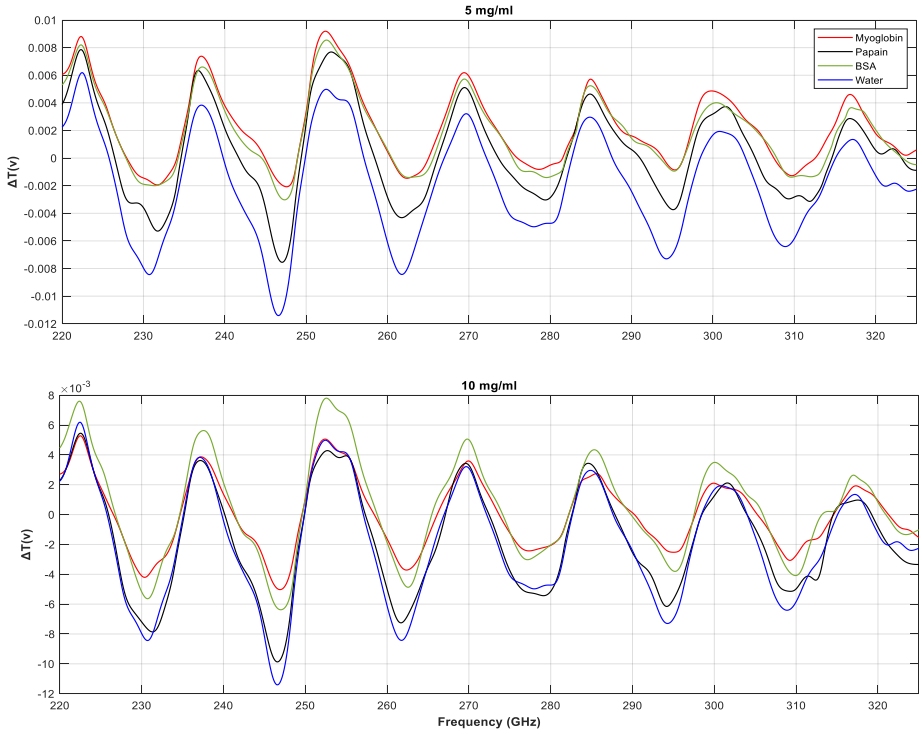

Figure 5.8: Sub-THz CD spectra of 5 mg/ml (upper) and 10 mg/ml (bottom) myoglobin, papain and bovine serum albumin solutions with pathlength of 500 µm. In the group of low concentration, myoglobin and bovine serum albumin solutions have no observed frequency shift at zero-crossing points indicating that the absorption of protein's CD of both samples are too weak to compare the differences under the strong water attenuation caused by longer pathlength. By increasing the concentration level, the CD spectrum of myoglobin and bovine serum albumin still have no observed frequency shifts, and CD spectra of papain are overlapping with that of water, also indicating that the water attenuation are significantly weakens the CD signal traveling in the sample solutions. The CD signals after samples is too weak to provide accurate a CD spectrum for identification.

In Fig. 5.8 the sample thickness is 500 µm. At both levels of concentration, MB and BSA have similar spectra and most of the zero-crossing points show no shift. With increasing concentration, CD spectra of the three proteins approach that for water. Comparing the upper plots in Fig. 5.6 and 5.8, the patterns are similar. CD spectrum of BSA and MB solution have no observed frequency shifts at each zero-crossing points indicate that low concentrated protein provide weak CD absorption and CD signal after sample cannot identify the difference between two proteins.

Tables 5.1 and 5.2 below show, respectively, zero-crossing frequency values for 10 mg/ml MB, Papain, BSA, and water for sample thicknesses of 100 and 310 µm. Although

BSA contains a relatively higher percentage of α-helices, CD from both samples has a high degree of similarity. An increase in sample thickness is needed to distinguish their CD. BSA, containing almost twice as many α-helices than Papain, has large frequency-shifts at the 9^{th} (1.38 GHz) and 13^{th} (1.15 GHz) zero-crossings (Table 5.1). However, both proteins contain a similar percentage of β-sheets, resulting in no observable shift (> 1 GHz) at other points. MB, however, shows frequency shifts at the 5^{th} (1.48 GHz), 7^{th} (1.46 GHz), 9^{th} (1.83 GHz), 11^{th} (2.17 GHz), and 13^{th} (2 GHz) zero-crossing comparing with Papain. At the 5^{th}, 7^{th}, and 11^{th} crossings, frequency shift of 1.3 GHz, 2.1 GHz, and 2.17 GHz occur between MB and BSA, while almost zero shift occurs for Papain and BSA. This indicates that β-sheets plays a significant role there. At the 7^{th} crossing, a frequency shift of 0.45 GHz between Papain and BSA suggest that α-helices still influence the outcome to a certain extent. Moreover, over 50% of α-helices result in the 9^{th} crossing becoming blue shifted by 1.83 GHz between MB and Papain, and 1.38 GHz between BSA and Papain. Analysis for the 100 μm pathlength shows α-helices strongly influence the 9^{th}, and 13^{th} crossings while β-sheets control the 5^{th}, 7^{th}, and 11^{th}events.

From Fig. 5.7, the 310 μm gap provides an ideal sample thickness for protein CD. Nine zero-crossing points report at least a 0.7 % (1.85 GHz between MB and Papain), frequency shift (red numbers in Table 5.2). The 2^{nd}, 4^{th}, and 8^{th} zero-crossing points (blue numbers) show frequency shift less than 0.5 % between proteins, 0.4% at 0.99 GHz, 0.2 % at 0.74 GHz, and 0.3% at 0.92 GHz, respectively. Although positions of zero-crossing points reports almost no shift between BSA and water in Table 5.2, the amplitudes of spectra indicate the existence of BSA in Fig. 5.7. MB of this thickness provides strong CD, indicating that the ratio of total α-helices significantly affects CD and diminishes the

positive reading of amplitude (amplitude above zero). By comparing both Table 5.1 and 5.2, α-helices and β-sheets influence CD spectra more efficiently for the 310 μm pathlength and are effective in inferring protein secondary structure. Since BSA contains over 50% α-helices and has a similar β-sheets-to-Papain ratio, its spectrum shows that all zero-crossing points lie between MB and BSA. To compare BSA with the 'standard' α-helical protein, MB, the influence of β-sheets over CD spectra should be analysed, because both proteins contain close ratios of α-helices relative to Papain. From Table 5.2, six out of ten zero-crossing points report larger (>1 GHz) frequency shifts between MB and BSA, namely at the 3rd (1.35 GHz), 5th (1.27 GHz), 7th (1.25 GHz), 9th (2.4 GHz), 11th (1.65 GHz), and 13th (1.8 GHz), respectively. To compare BSA and Papain, it is discovered that α-helices strongly influence the 1st (1.75 GHz), 3rd (2.18 GHz), 5th (2.25 GHz), 6th (1.39 GHz), 7th (2.09 GHz), 9th (2.99 GHz), 10th (1.14 GHz), 11th (2.78 GHz), and 13th (2.2 GHz) zero crossings. When the ratio of α-helices increases, the odd numbers of zero-crossing points are red shift while even numbers of points are blue shifted, especially above 280 GHz. A higher ratio of α-helix suffers stronger CD absorption and pulls zero-crossing points closer to each other (8th and 9th, 10th and 11th, and 12th and 13th). From the analysis above, the sub-THz CD spectrum is strongly dominated by α-helix. β-sheet, on the other hand, has less influence over the spectrum except at the 9th, 11th, and 13th zero-crossing points.

Comparing both the 100 and 310 μm pathlengths, more numbers of zero-crossing points with frequency shift are observed between proteins for the 310 μm pathlength. CD spectra for a 100 μm pathlength provide more accurate information of how β-sheets influence CD spectra, registers no shifts at the 5th and 11th zero-crossing points. By increasing the pathlength, high ratios of α-helices provide much stronger CD absorption than that of β-

sheets. For the 100 μm case, the influence of β-sheets at the 5th and 11th zero-crossing points is clearly identified, as no observable shift occurs between Papain and BSA. Due to the ratios of α-helices and β-sheets at those three points, it is significant to note how secondary structure affects sub-THz CD spectra in both pathlengths.

Table 5.1: Frequency values of circular dichroism zero-crossing points for a 10 mg/ml sample of protein with 100 μm signal propagation depth. Red numbers represent the zero-crossing points where frequency shifts are observed. Myoglobin has a higher ratio of α-helices, reporting frequency shifts at the 5th, 7th, 11th, and 13th zero-crossing points (> 1 GHz). Papain and bovine serum albumin both contain β-sheets, thus no shifting is observed except at 9th point (1.38 GHz frequency shift).

Zero-Crossing α-helices β-sheets	Myoglobin (GHz) 75% 0	Papain (GHz) 25% 21%	BSA (GHz) 54% 18%	Water (GHz)
1st	226.55	225.59	225.52	227.51
2nd	234.64	235.05	234.91	234.39
3rd	241.55	240.20	240.50	243.20
4th	249.54	249.97	249.94	249.36
5th	258.78	257.30	257.48	260.03
6th	266.30	266.55	266.19	265.43
7th	274.00	272.54	272.10	274.56
8th	282.23	282.70	282.70	282.00
9th	289.83	288.00	289.38	292.35
10th	297.18	298.00	297.83	296.74
11th	305.94	303.77	303.77	307.09
12th	314.30	315.24	315.36	313.85
13th	321.50	319.50	318.35	322.56

Table 5.2: Frequency values of circular dichroism zero-crossing points for a 10 mg/ml sample of protein with 310 μm signal depth. Red numbers represent the zero-crossing points where frequency shifts are observed (> 1 GHz). When comparing with bovine serum albumin, myoglobin has higher ratios of α-helices, resulting in frequency shifts at 3rd, 5th, 7th, 9th, 11th, and 13th zero-crossing points. Papain and BSA both contain β-sheets with observable frequency shifts at 1st, 3rd, 5th, 6th, 7th, 9th, 10th, 11th, and 13th zero-crossing points.

Zero-Crossing α-helices β-sheets	Myoglobin (GHz) 75% 0	Papain (GHz) 25% 21%	BSA (GHz) 54% 18%	Water (GHz)
1st	225.63	227.94	226.19	226.45
2nd	235.18	234.19	234.80	234.82
3rd	239.80	243.33	241.15	241.44
4th	249.83	249.09	249.50	249.65
5th	257.00	260.52	258.27	258.38
6th	267.02	265.16	266.55	265.74
7th	272.06	275.45	273.31	273.45
8th	282.72	281.80	282.30	282.40
9th	287.50	292.89	289.90	290.20
10th	298.38	296.41	297.55	297.44
11th	303.15	307.63	304.85	305.00
12th	315.18	313.84	314.47	314.27
13th	318.55	322.55	320.35	319.90

5.4 Summary

This chapter reports seminal sub-THz circular dichroism measurements of myoglobin (MB), bovine serum albumin (BSA), and papain from 220 to 325 GHz. Each was dissolved in pure water. The secondary structure of MB contains eight folded α-helical segments and so provides a good 'standard' in the study of sub-THz CD. Another two proteins, BSA and Papain, containing different ratios of α-helices and β-sheets to MB, are measure for CD in comparison. With sample thickness and concentration varied, the α-

helices provided CD. Sample thicknesses of 500 μm proved too great for the present capability of the apparatus to observe CD; the amount of associated water absorbing all the signal. By analysing CD spectra between proteins, the way α-helix and β-sheet control frequency shifts at each zero-crossing points can be predicted. In a 100 μm pathlength, spectra show observed frequency shift at five zero-crossing points by comparing three proteins. However, by comparing BSA and Papain containing close ratios of β-sheets, no shift is observed at the 5^{th}, 7^{th}, and 11^{th} points other than a large (1.38 GHz) frequency shift at the 9^{th} point. Since MB contains only α-helical segments, the frequency shifts at the 5^{th}, 7^{th}, and 11^{th} points are presumably mainly caused by β-sheets. The 9^{th} and 13^{th} zero-crossing points on the other hands, are dominated by α-helices. By increasing the pathlength to 310 μm, ten out of thirteen points report frequency shifts between proteins. By comparing three proteins, frequency shifts are observed at 3^{rd} (1.35 GHz), 5^{th} (1.27 GHz), 7^{th} (1.25 GHz), 9^{th} (2.4 GHz), 11^{th} (1.65 GHz), and 13^{th} (1.8 GHz) between MB and BSA. However, by comparing BSA and Papain, even larger frequency shifts are observed at 1^{st} (1.75 GHz), 3^{rd} (2.18 GHz), 5^{th} (2.25 GHz), 6^{th} (1.39 GHz), 7^{th} (2.09 GHz), 9^{th} (2.99 GHz), 10^{th} (1.14 GHz), 11^{th} (2.78 GHz), and 13^{th} (2.2 GHz) points due to the effect of α-helices. By comparing both results, α-helices dictate the positions of nine zero-crossing points as above. β-sheets, on the other hand, show less influence on the CD spectrum. Strong water attenuation and absorption from high ratios of α-helices, render the presence of β-sheets nearly invisible. From the measurements, spectra in the group of 100 μm sample pathlength show a clear influence at each zero-crossing point from α-helices and β-sheets among the different proteins. Such a phenomenon is valuable in the study of protein secondary structure. However, the greater the spectral-shift in zero-crossing, the higher the

accuracy in identifying the sub-THz CD responses corresponding to the protein. The CD spectrum of proteins for the 310 μm pathlength provide more zero-crossing points and a larger frequency shift (>1 GHz), giving it a clear advantage in identifying different proteins based on their ratios of α-helices. To avoid the influence of α-helices, comparing two proteins of the same ratio of α-helices but not β-sheets are required. However, such approach would require further investigations.

Chapter 6

Dielectric Study of Solvated Proteins Using Terahertz Time-Domain Spectroscopy

6.1 Introduction

Solvated Myoglobin (MB), bovine serum albumin (BSA), and Papain have been studied using the THz QO CD system in Chapter 5. Different ratios of α-helix and β-sheet in the secondary structure of the protein influences the THz CD spectra. According to the findings in Chapter 5, CD spectra of three proteins in the group with 100 μm sample thickness indicates that MB provides more CD absorption due to its pure α-helical protein (75%). BSA, containing a lower ratio of α-helixes (54%), thus providing less CD absorption than MB while exhibiting more frequency shifts between zero-crossing points than Papain. Hydrogen bonds, covalent bonds, and hydrophobic interactions also play their critical roles in the formation of protein structures [101]. Such phenomena results in unique dielectric properties for each type of protein. In recent years, dielectric study of protein has shown a growing interest in [30, 101-105]. The sub-THz CD QO system is a new technique in the research of protein features, and its dielectric study of protein can help analyse the protein's behaviour in sub-THz spectral domain.

The many-fold physiological functions of solvated proteins are determined by conformational shapes they adopt under the driving forces exerted between hydrophilic and hydrophobic portions of the molecule, and water. Hydration bonding, therefore, plays a vital role in stabilizing protein secondary structure and governance of their dielectric properties [68, 105, 106]. Since biological polymers such as proteins have intrinsic,

permanent polar character; probing electromagnetic radiation may influence the dielectric character of proteins [68]. Fig. 6.1 shows four types of polarization influencing the dielectric spectrum over a wide domain of frequencies, namely interfacial ($> 10^6$ Hz), orientational ($10^6 - 10^{10}$ Hz), atomic (0.01 to 10 THz), and electronic (< 10 THz) [105].

Electromagnetic radiation of millimetre and sub-millimetre wavelength are resonant with chain or backbone vibrational modes of a protein [26, 30, 108]; besides being able to couple to the relaxation dynamics of water [27, 28, 109]. According to Adalberto *et al* [104], protein solutions exhibit dielectric water relaxation in the radio frequency domain, governed by orientation polarization. THz time-domain spectroscopy (THz-TDS) has been applied over the past few decades to extract the complex refractive index of solvated proteins [11, 68, 102]. For this, advanced signal processing methods using THz-TDS are required to record more accurate material properties of solvated proteins [11, 102, 110]. The Fabry-Perot effect is usually taken into consideration when studying multiple-layer samples. Sushko *et al* [102] have enhanced the accuracy of extracting the complex refractive index of materials by factoring etalon it in.

In this chapter MB, BSA and Papain, were chosen as systems to correlate secondary structure with dielectric and CD response. To reiterate from earlier, the secondary structure of MB consists of eight α-helices (75%) with no β-form; BSA contains approximately 54% α-helices and 40% β-form; and Papain, 25% of α-helices and 21% of β-sheets. Since protein conformations must influence dielectric behaviour, the three proteins are expected to exhibit distinct dielectric response in the sub-THz spectral domain. An aim now is to correlate dielectric study with sub-THz CD for mutual instruction about protein response to environmental alterations in its host medium, e.g. temperature, pH, salinity etc.

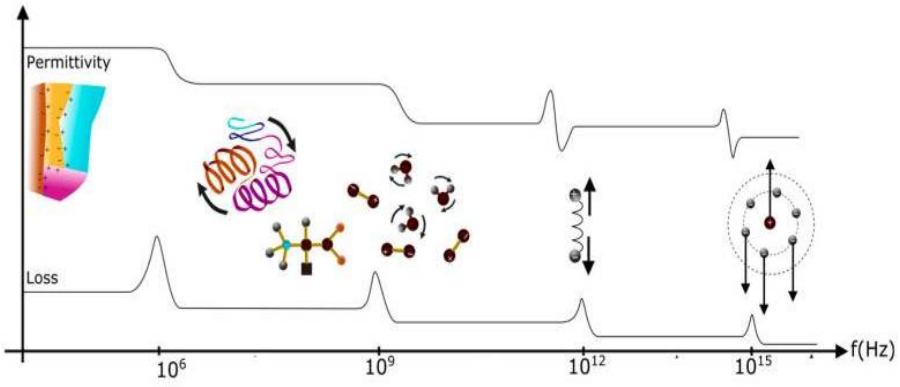

Figure 6.1: The real (top trace) and imaginary parts of permittivity as a function of frequency [105]. The dielectric response can be grouped into five categories of charge character: polarisation, interfacial ($> 10^6$ Hz), orientational (dipole) ($10^6 - 10^{10}$ Hz), atomic (0.01 to 10 THz), and electronic (< 10 THz). The sub-THz to THz domain spans 10^9 to 10^{12} Hz, coinciding with molecular vibrational modes of motion. From DC to approximately 3 GHz, the relative permittivity of water is approximately 79 at 20℃ [111]. At optical frequencies, it drops to approximately 2 due to the small net polarisation [112]. The permittivity of liquids broadly decreases with increasing temperature increases [111].

6.2 Dielectric Property extraction by THz-TDS

The properties of protein solutions in THz-TDS measurement is determined by the comparison of two signals in the frequency domain, a sample signal $\vec{E}_s(\nu)$ and a reference signal $\vec{E}_{ref}(\nu)$. The reference signal is obtained by taking a measurement with an empty sample holder. A sample signal is measured in this case by introducing a liquid sample into the empty cell. The transfer function of the dissolved protein sample is given by the ratio [102, 113]

$$H(\nu) = \frac{\vec{E}_s(\nu)}{\vec{E}_{ref}(\nu)}. \tag{6.1}$$

The Fresnel equations for the case of normal incidence, describe transmission t_{ij} (from medium j into medium i) and reflection r_{ij} (in medium i from the j^{th} material layer). The transmission coefficient for the amplitude of the emerging electric fields with respect to the incident is:

$$t_{ij} = \frac{2n_i}{n_i + n_j}$$

$$r_{ij} = \frac{n_j - n_i}{n_i + n_j} \tag{6.2}$$

n is the refractive index. Thus, the propagation factor $\vec{P}(\nu)$ describes an electromagnetic field propagating through a material of thickness d with its transfer function expressed as [112, 113]:

$$\vec{P}(\nu) = e^{(\frac{-i2\pi\nu\tilde{n}(\nu)d}{c})}$$

$$\overrightarrow{H_s}(\nu) = t_{12} \times t_{21} \times \vec{P}(\nu) \tag{6.3}$$

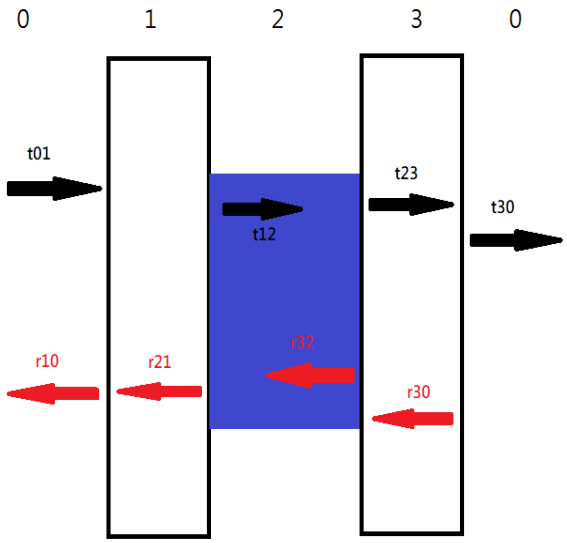

Figure 6.2: Schematic of reflection and transmission at the interface of sample and TPX windows of the cuvette. Black arrows represent transmission, and red represent reflection. Material 1 and 3 are TPX windows. t_{01} is the incident wave before propagating through the TPX. t_{12} is transmission in the sample. r_{21} and r_{32} are the reflection reflected by the surface between the sample and the TPX. r_{10} and r_{30} are the reflective wave reflected by the surface between the TPX and air.

A Bruker cell was used as a cuvette. It consists of an outer pair of TPX windows, an inner pair of Teflon gaskets, and a central spacer. The spacer enforces a rigid gap of required

size between the TPX windows and allows the liquid sample to be injected into the cuvette. The cuvette can be considered as a three-layer system. Fig. 6.2 is a schematic of this showing the primary and subsequent sequential orders of internal refraction and transmission. An accurate estimate of the dielectric function of the sample depends upon correctly factoring in the Fabry-Perot (FP) or 'etalon' effect [102, 113] and is given by the following expression [102]:

$$FP(\nu) = \sum_{l=0}^{m} [\vec{r}(\nu)^2 \vec{P}(\nu)^2]^l. \tag{6.4}$$

Thus, the theoretical transfer function of a protein solution as derived from (6.3) and (6.4) is [102]:

$$\overrightarrow{H_s}(\nu) = (t_{01} \times t_{12} \times t_{23} \times t_{30}) \left(\vec{P}_{air}(\nu) \times \vec{P}_{window}(\nu) \times \vec{P}_{sample}(\nu)\right)$$

$$* [1 + (2 \times r_{32} r_{30} e^{\left(\frac{-i2\pi\nu\tilde{n}_w(\nu)2d_w}{c}\right)}) + (2 \times r_{32} r_{30} e^{\left(\frac{-i2\pi\nu\tilde{n}_w(\nu)2d_w}{c}\right)})^2] \tag{6.5}$$

where $r_{30} = \frac{n_{air} - n_w}{n_w + n_{air}}$ is the reflection coefficient at the interface of the TPX window and air; and $r_{32} = \frac{n_{sample} - n_w}{n_w + n_{sample}}$ that for the liquid sample and TPX window; d_w is the thickness of the TPX windows measured to be 1.71(5) mm. The thickness of the sample solution is 100 μm. The refractive index of TPX is 1.46 [102]. Since TPX windows are extremely low-loss, its extinction coefficient $k(\nu) \cong 0$. Therefore, by matching the measured and theoretical transfer functions via a non-linear regression analysis, such that $\vec{H}(\nu) = \overrightarrow{H_s}(\nu)$, the refractive index of the sample can be calculated. Fig. 6.3 shows the refractive index and absorption coefficient of water. The refractive index drops rapidly in value for

frequencies below 1 THz. The refractive index of water at 1 THz is 2.29 (error of 2.6% comparing with approximately 2.35 [114] and error of 1.74 % comparing with approximately 2.25 [115]). At H band (220 to 325 GHz), the refractive index dips from 3 to 2.73. Differences of 3.3% (Wilmink et al [114] who measure from approximately 2.9 to 2.7) and 8.3% (Ronne et al [115] who measure approximately between 2.75 to 2.5) respectively are due to the different methods of extraction. In most dielectric studies, multiple reflections are minimal and Fabry-Perot effects are neglected due to the high absorption of water [116]. However, in our case, reflections and Fabry-Perot effects will be considered to obtain more accurate data. Both the real and imaginary parts of dielectric permittivity of protein solution are calculated with (6.1) and presented in Fig. 6.4.

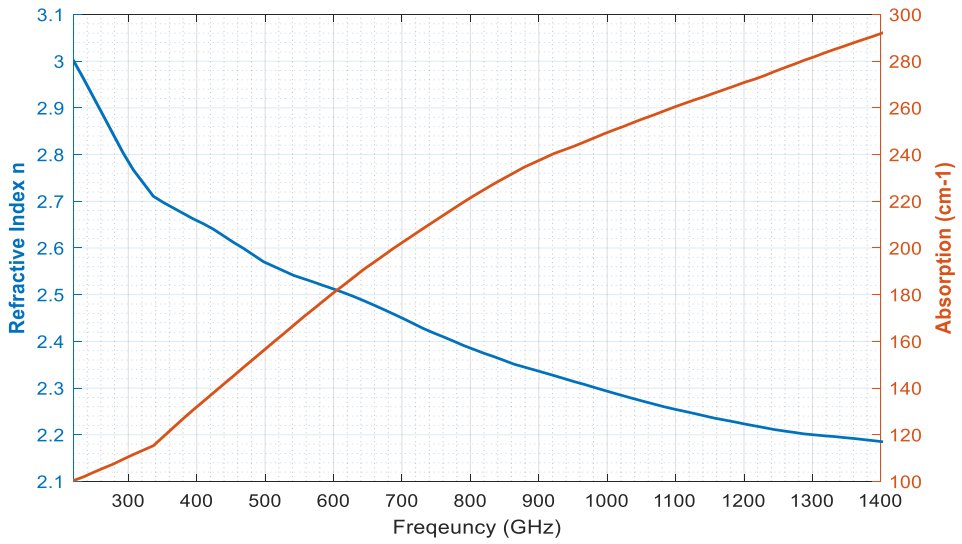

Figure 6.3: Refractive index (blue) and absorption coefficient (red) of water from 200 to 1.4 THz using THz-TDS. The values of refractive index are in the range of 3 to 2.18 from 200 GHz to 1.4 THz. The values of absorption coefficient lie in the range of 100 to 295 cm^{-1}. The absorption coefficient, A, is related to the extinction coefficient by $A = k(v) \times \frac{4\pi f}{c}$. The extinction coefficient is a measure of the component of signal that does not couple in-phase with a system it is interacting with. The absorption coefficient is a measure the distance traveled by the electromagnetic wave inside the sample to the point where its amplitude is diminished to a value of 1/e of its initial value. This is an effective distinguishing measure of materials and is used to compare with THz CD results of Chapter 5.

6.3 Results

Table 6.1: Protein properties which are calculated by the protein dipole server [117]. PDB files were chosen from the RCSB Protein Data Bank.

	Numbers of residues	No. of positive residues	No. of positive residues	Total protein net charge (e)	protein dipole moment (D)
Myoglobin PDB ID:1MBN	**153**	**23**	**21**	**2**	**239**
BSA PDB ID: 3V03	**581**	**82**	**98**	**-16**	**579**
Papain PDB ID: 9PAP	**211**	**22**	**13**	**9**	**226**

A dielectric analysis of the protein solutions is sought to be understood in connection with their sub-THz CD activity. Myoglobin, BSA, and Papain were noted above to consist of different ratios of α-helices and β-sheet. From Chapter 5 is was seen that sub-THz CD was sensitive to varying secondary structure in response to changes in pH, concentration etc. Different dielectric behaviour is similarly expected for different protein conformation. The sample thickness was set at 100 μm to avoid strong attenuation during TDS measurements. The concentration of the protein solutions is each 10 mg/ml. To ensure sound correlation between THz CD and dielectric measurement, all the protein samples in Chapter 5 and 6 were prepared simultaneously.

Fig 6.4 co-plots for comparison the dispersive refractive index and absorption coefficient of MB, BSA, and Papain at D band. Fig. 6.5 extracts the real and imaginary parts of the dielectric permittivity. In Fig. 6.4, Papain has a lower absorption coefficient for its lower sub-THz CD sensitivity. According to Sushko et al [37], the hydrophobic residues are negatively correlated with the absorption coefficient. The estimated ratio between hydrophilic and hydrophobic surfaces for MB and BSA are 1.31 and 1.51, respectively [37]. BSA has a higher absorption coefficient due to the highest total hydrophilic area that nearly matches [37]. Since the majority of Papain's residues are

hydrophobic, such finding explain the lower absorption coefficient. Furthermore, protein dipole moment influences the absorption coefficient as well. In Table 1, the total protein dipole moment of MB, Papain, and BSA are 239, 226, and 579 D. The protein dipole moment alters water dynamics around the protein surface. According to Zhang et al [97], the dipole moment of water molecule reduces from 2.5 D in the bulk water state to approximately 0.8 D water molecule around the protein, exhibiting a decrease in absorption. However, the absorption of CD is only sensitive to the chiral structure of α-helices and β-forms demonstrated in Fig. 5.6 and 5.7. Although there is no research that explores the correlation between protein absorption coefficient and the absorption in protein CD, by comparing the protein absorption coefficient and THz CD spectra in Fig. 5.6, 5.7, 5.8, it becomes evident that proteins provide higher CD absorption and cause an observable shifts between zero-crossing points above 260 GHz frequency as the absorption coefficient increases. Based on this observation, proteins with higher absorption coefficients are providing stronger CD protein absorption in the THz CD measurement. This phenomenon also helps explain why both the amplitude of Papain sub-THz CD and frequency zero-crossing points show no observable shifts or changes when compared with water in Fig. 5.8. Under strong water attenuation, solvated Papain exhibits almost no sub-THz CD. For MB and BSA, only weak CD is detected due to a higher protein absorption coefficient. From the absorption coefficient plots of Fig. 6.4, CD is more active at higher frequency. Furthermore, the traces of the protein absorption coefficient match with the sub-THz CD spectra at lower concentration level (Fig.5.6 and 5.7; 5mg/ml) and longer path-length (Fig. 5.8; 5 mg/ml and 10 mg/ml). Unlike sub-THz CD QO system, THz-TDS is extremely sensitive to the water due to the speed of water molecule rotational motion [30, 37]. The

absorption coefficient rapidly drops once the concentration level rises up to 50 mg/ml [37]. This is because the total volume of hydration shells increases with protein concentration at low concentration levels. However, in THz CD measurement, the concentration level increase would lead to an increase in the pathlength, providing CD absorption between co- and cross-polar signals by α-helices and β-forms.

In comparing their respective secondary structures, MB and BSA contain a higher ratio of α-helices and β-form (over 70 %). Papain, by contrast, has less than 50%; so a basic correlation is observed between secondary structure and dielectric response. Due to the electronegativity of oxygen and the distance between oxygen and hydration atoms, water molecules have a dipole moment in the range between 2.4 to 3.0 D [97]. This helps explain why water shows high permittivity at low frequency. However, protein molecules usually have a rotational relaxation time of picoseconds (BSA has a relaxation time of approximately 9.5 ps at 16 GHz [118]). This means that protein dipole moment yields strong response to the driving radiation at lower frequency than in the sub-THz frequency band as describe in Fig. 6.1. However, the protein net charge and hydrophilic side chains of amino acids couple strongly with the probe EM radiation between 220 to 325 GHz [37]. The net charge and numbers of charged residues of proteins are shown in Table 1. The relative permittivity in Fig. 6.5 suggests that solvated Papain has higher relative permittivity. BSA, which contains more negative residues than positive, has lower permittivity. However, BSA has a total protein dipole moment of 579 D, almost 2.5 time more than Papain and MB. It also indicates that protein dipole moment produces a weak response in the sub-THz frequency band. In pure water, hydration bonds easily break and reform; not so in association with proteins. Hydration bonds inside α-helices and β-sheets

can be formed by the carbonyl oxygen bonds and amino group on the main chain [119]. Side chains, containing oxygen and hydration atoms, may form hydration bonds as well. Such hydration bonds inside the protein are crucial for forming altered secondary structures in the folding states. Furthermore, the hydrophilic and hydrophobic effects of protein can strongly influence the contact between the protein surface and surrounding water and impact protein−ligand binding [118]. Hydrogen bonds, charged residues, and hydrophilic and hydrophobic effects of protein are the main contributing factors in electrostatic interaction critical for the protein structure, resulting in unique responses of dielectric permittivity.

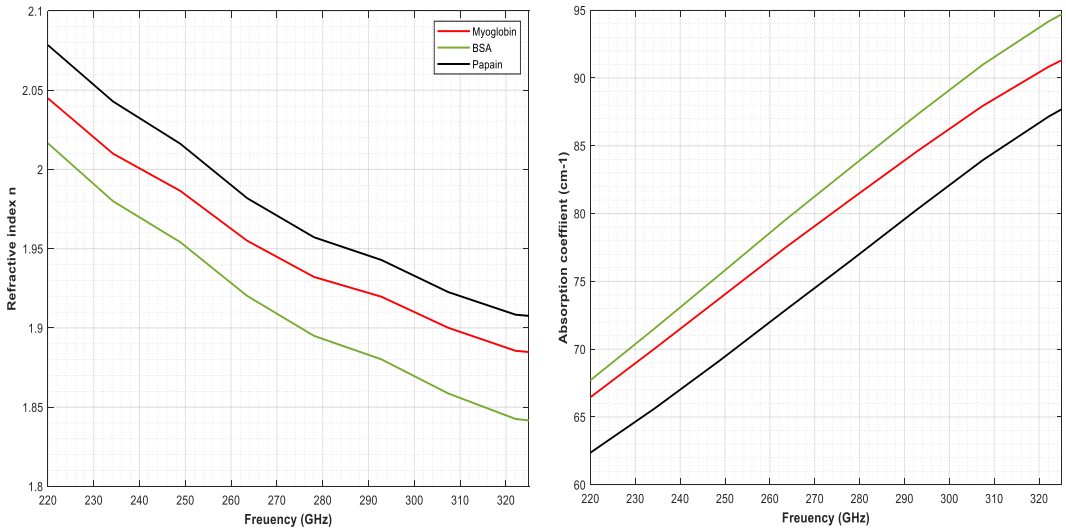

Figure 6.4: Refractive Index (left) and Absorption coefficient (right) of solvated MB (red 66.5 to 91.1 cm^{-1}), BSA (green 68 to 94.8 cm^{-1}), and Papain (black 62.5 to 87.7 cm^{-1}) with 100 µm thickness from 220 to 325 GHz. The protein absorption coefficient will be used for comparison with the THz CD spectra in Chapter 5. Since THz TDS is extremely sensitive to water, it is anticipated that the absorption coefficient of protein can be used to analyse the THz CD spectra of proteins with sample thicknesses of 310µm (Fig. 5.7).

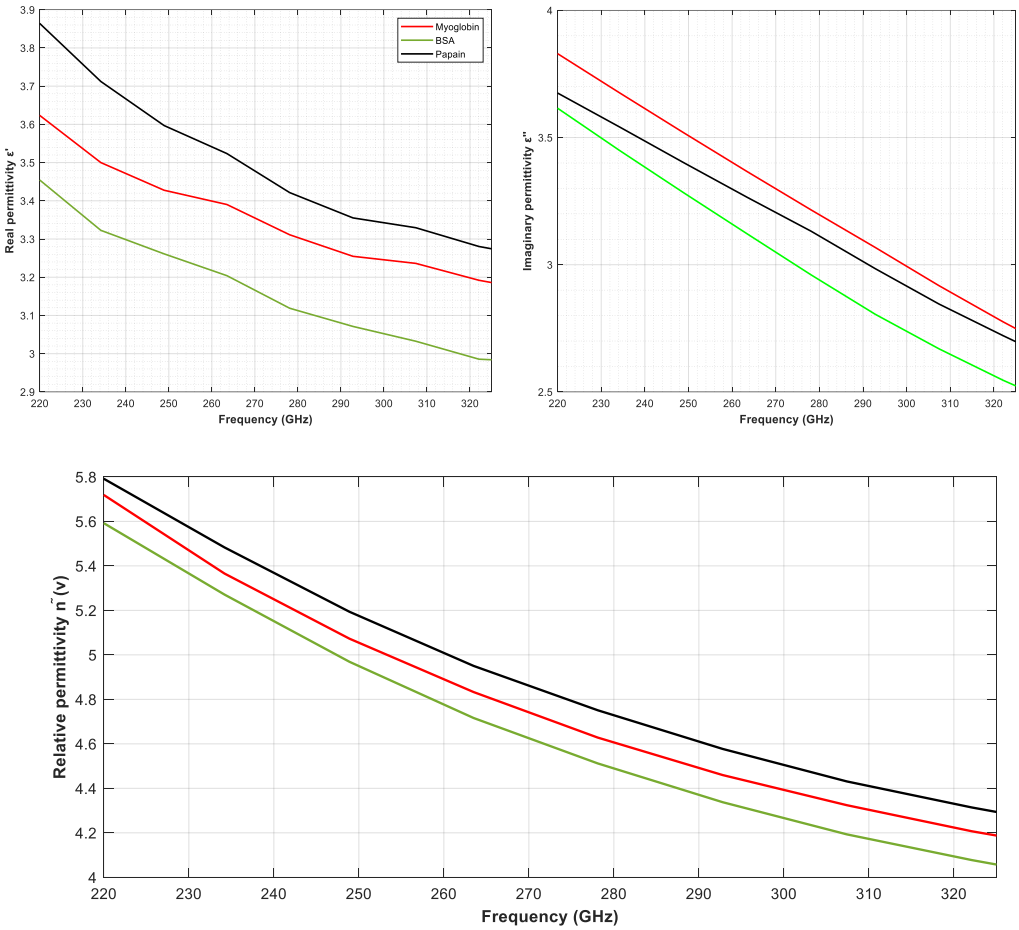

Figure 6.5: The real (upper left) and imaginary (upper right) parts of the dielectric permittivity of the proteins from 220 to 325 GHz. The bottom plot shows the relative permittivity $\tilde{n}(\nu)$ of the proteins considered. The protein dipole moment provides weak response to sub-THz radiation due to the slow protein relaxation time. The protein net charge and hydrophilic side chains of amino acids interact stronger with the incident radiation in this band.

6.4 Summary

The Myoglobin (MB), bovine serum albumin (BSA), and Papain solutions from Chapter 5 have been used for the study of their dielectric properties using THz-TDS from 0.2 – 1.0 THz. The three proteins, having differing ratios of α-helices and β-structures in their secondary structures, and show distinctly different dielectric responses. Electrostatic forces in action between water and protein are critical in the stability of protein secondary structure. This is mainly attributed to hydrogen bonds, hydrophobic and hydrophilic

surfaces, and positively and negatively charged residues. Due to a higher ratio between hydrophilic to hydrophobic surfaces, BSA has a higher absorption coefficient than MB and Papain. By comparing the absorption coefficient with sub-THz CD spectra, the traces indicate that a higher absorption coefficient of protein correlated with a more sensitive CD response when frequencies increasing. For the dielectric permittivity, a protein dipole moment only responds weakly to THz radiation since the protein relaxation time is in order of picoseconds (between 10 to 30 GHz). In sub-THz range, hydrophobic and hydrophilic surfaces, and protein net charge couple more efficiently to the incident radiation. Unlike the absorption coefficient and dielectric permittivity of protein, the sub-THz CD absorption only responds to certain ratios of α-helices and β-structures. Since THz TDS is extremely sensitive to water, protein concentration and sample thickness significantly influences the measured dielectric response. Several research findings have demonstrated that protein concentration alters the hydrogen shell around the protein. The protein absorption coefficients remain constant as the concentration reaches 50 mg/ml [27]. For lower concentrations, a high percentage of water enhances the water molecule dynamics around the protein surface and protein rotational dynamics. A protein surrounded by water molecules gains more rotational freedom and so provides a strong THz response, resulting in a high absorption coefficient. Based on current findings between protein dielectric properties and sub-THz CD, a dielectric study of protein helps explore how the electrostatic interaction influences α-helices and β-structures in the protein structure in various conformational states, as well as the effects of varying ratios between α-helices and β-structures on THz CD.

Chapter 7

Conclusion

A new experimental methodology has been developed and applied to measure the circular dichroism of selected solvated proteins over a continuous span of terahertz frequencies (principally 220 – 320 GHz). The continuous nature of measurement was achieved by a modern, high-performance vector network analyser interfaced to a quasi-optical circuit. The THz CD spectrum of the solvated protein is correlated with protein secondary structure which has been demonstrated in Chapters 4 and 5. THz time-domain spectroscopy has been applied to measure the dielectric properties of selected proteins and a correlation to THz Circular Dichroism (CD) has been sort.

The contributions in this thesis are:

- With the aid of Jones' calculus and vector network analysis, two measuring schemes of sub-THz Circular Dichroism Quasi-Optical System were demonstrated. The first involved the rotation of the receiving horn to directly measure the cross-polar component of transmission. This invariably introducing significant error as moved cabling results in overall departure of the network from initial calibration conditions. Scheme 2 therefore introduced a methodology that preserved the condition of the calibration of the network. Solvated Myoglobin was measured on three separate occasions to test measurment system stability and repeatability.
    - The protein concentration level can influence the performance of the sub-THz CD spectra. At the D mm-waveband, the pure buffer suffered more phase noise due to strong, sensitive absorption which can influence the performance of the measurement. At H mm-waveband the sub-THz CD spectra for 1.00 mg/ml and

0.25 mg/ml MB samples are similar to that for the pure buffers, and exhibit statistically significant frequency (> 1 GHz) at each zero-crossing point. With the concentration level raised to 5 mg/ml, Myoglobin solutions presented 7 zero-crossings at D mm-waveband and 13 at H mm-waveband.

- The demonstration that the pH values can influence the secondary structure of protein and cause the frequency shifts between zero-crossing points. In the far-UV CD study, the pH values of 8.26 and 5.50 reduce the CD reading by influencing hydrogen bonds inside the folding MB. In sub-THz CD study, the pH 5.50 buffer was measured to have a stronger influence over the MB secondary structure. At H mm-waveband, such an influence results in observed frequency shifts (> 1 GHz) at the 2^{nd}, 5^{th}, 8^{th}, 9^{th}, 10^{th}, and 11^{th} zero-crossing points.
- The different secondary structure of proteins contain different ratios of α-helices and β-forms that can provide different sub-THz protein CD spectra. By comparing the spectra of MB, BSA, and Papain, α-helices with a 100 μm path-length strongly influence 9^{th}, and 13^{th} points, while β-sheets influence 5^{th}, 7^{th}, and 11^{th} points. In the group of 310 μm sample thicknesses, it is discovered that α-helices strongly influence 1^{st}, 3^{rd}, 5^{th}, 6^{th}, 7^{th}, 9^{th}, 10^{th}, 11^{th}, and 13^{th} (frequency shift at each points > 1.46 GHz) by comparing BSA and Papain, since both proteins contain similar ratios of β-sheets. Papain contains only 25 % α-helices, rendering it less sensitive to CD signals. Under strong water attenuation (500 μm), the sub-THz CD spectrum of Papain is much closer to that of water with no frequency shifts observed at zero-crossing points. However, Myoglobin and

BSA (that contain higher ratios of α-helices), have no observed frequency shift when compared to each other. From the analysis, α-helices provides stronger sub-THz CD responses than β-form and dominates the overall sub-THz CD spectra.

- By analysing the dielectric properties of proteins, the possibility of a physical link between sub-THz CD spectra and protein structures was explored. From the analysis, the absorption coefficient of BSA is higher than those of MB (3.94% at 325 GHz) and Papain (7.97 % at 325 GHz) in the H mm-waveband due to the higher ratio of hydrophilic to hydrophobic surfaces and higher protein dipole moment. The estimated ratios of MB and BSA are 1.31 and 1.51, respectively. Since most residues in Papain are hydrophobic, it is expected that Papain has a lower absorption coefficient in this band. Based on the analysis in Chapter 5 and 6, proteins provide higher CD absorption and cause an observable shift between zero-crossing points above 260 GHz frequency as the absorption coefficient increases.

THz CD is a new tool complementing established biological methodologies and is operable in the liquid state. Quasi-Optical sub-THz CD has successfully distinguished between three proteins that differ in their ratios of α-helix to β-form. However, protein structure is too complex in the THz domain of frequencies (supporting thousands of rotational and vibrational modes of motion). To understand how the α-helix and β-form are manifested in the THz CD spectrum, simpler structures of the bio-molecule such as peptide, should be chosen, using the methodology presented in Chapter 4 and 5 in the future work. Polypeptide contains a single linear chain which supports only few modes in the THz

domain of this thesis. Such characteristics can be captured in the form of a THz CD signal easily, avoiding noise from random motions. With regard to the sample holder, after normalizing the background signal and the effect of TPX windows, the sample holder was moved from its original position to inject liquid sample. However, this will introduce small yet significant manual error. To avoid this, future work should include improvement of the sample holder. A micro-fluid system is being adopted to this end. THz CD spectroscopy successfully reduces the 'water effect' in the study of liquid samples. Water attenuation increases with increasing depth of sample. Heavy water (D_2O) is a good alternative. It is often used in acquiring FTIR spectra of protein in solution. Future work should include analysing these various improvements in using THz TDS and THz QO-CD systems and a database created. A quasi-optical sub-THz CD of proteins should now be preceded by a systematic and detailed study of constituent peptides and amino acids to elucidate their constituent role in the final character of a protein.

Many have studied the THz responses of biological samples using THz-TDS, and the theory and techniques of applying CD via sub-THz quasi-optics is transferable to THz-TDS. Room for improved performance lies too in using a large-area ferrite plate, in which only a small region at its centre is illuminated. This is alternate to the small tile used in this thesis. The uniformity and quality of the magnetization normal to the large face of the ferrite is degraded with reduced square-area (as was done here to meet practical constraints of space in building a quasi-optical circuit). The TDK hexaferrite plates, which serendipitously have a stoichiometry leading to low dielectric loss at sub-THz frequencies, should in future be retained in their large tile form. Poor magnetization performance can lead to absorption of upto 70% of the incident power with consequent degrading of the

detector performance. The reflection mode THz TDS has been used to study thin film [120] and liquids [121, 122] for years. This method provides for complementary investigation to improve the utility of THz CD spectroscopy on THz TDS in the future work.

Reference

[1] J. W. Fleming, “High resolution submillimeter-wave Fourier-transform spectrometry of gases,” IEEE Trans. Microwave Theory Tech., vol. MTT-22, pp. 1023–1025, 1974.

[2] Ghann, W. and Uddin, J. (2019). Terahertz (THz) Spectroscopy: A Cutting-Edge Technology.

[3] Peter H. Siegel “Terahertz Technology in Biology and Medicine”, IEEE, 2004.

[4] E. Pickwell, B. E. Cole, A. J. Fitzgerald, M. Pepper, and V. P. Wallace, “In vivo study of human skin using pulsed terahertz radiation”, Institute of Physics Publishing, Physics in Medicine and Biology, 2004

[5] F. Wahaia “Spectroscopic and Imaging Techniques Using Terahertz Frequency Band For Biomedical Applications,” PhD in Physics Thesis, University of Porto, 2011. 11

[6] Gun-Sik Park, Yong Hyup Kim, Haewook Han, Joon Koo Han, Jaewook Ahn, Joo-Hiuk Son, Woong-Yang Park, Young Uk Jeong, “Convergence of Terahertz Sciences in Biomedical Systems”, 2012.

[7] Yuko Ueno and Katsuhiro Ajito, “Analytical Terahertz spectroscopy,” *Analytical Sciences*, Vol. 24, pp 185-192, 2008.

[8] Jens Neu and Marco Rahm, “Terahertz time domain spectroscopy for carrier lifetime mapping in the picosecond to microsecond regime”, Optical Society of America, 2015

[9] K. Lien Nguyen, Tomislav Fri, Graeme M. Day, Lynn F. Gladden, and William Jones, “Terahertz time-domain spectroscopy and the quantitative monitoring of mechanochemical cocrystal formation”, Nature Material, 2007

[10] Trung Quan Luong, Pramod Kumar Verma, Rajib Kumar Mitra, and Martina Havenith, “Do Hydration Dynamics Follow the Structural Perturbation during Thermal Denaturation of a Protein: A Terahertz Absorption Study”, Biophysical Journal, 2011.

[11] Nicholas Y. Tan, Ruoyu Li, Pierre Brauer, Carmine D’Agostino, Lynn F. Gladden and J. Axel Zeitler, “Probing hydrogen-bonding in binary liquid mixtures with terahertz time-domain spectroscopy: a comparison of Debye and absorption analysis”, Royal Scoiety of Chemistry, 2015.

[12] Robert J. Falconer, and Andrea G. Markelz, “Terahertz Spectroscopic Analysis of Peptides and Proteins”, J Infrared Milli Terahz Waves, 2012

[13] Andrea G. Markelz, “Terahertz Dielectric Sensitivity to Biomolecular Structure and Function”, IEEE Journal of Selected Topics in Quantum Electronics, 2008.

[14] Vondracek H, Dielmann-Gessner J, Lubitz W, Knipp M, and Havenith M, “THz absorption spectroscopy of solvated β-lactoglobulin”, AIP, The journal of Chemical Physics, 2014.

[15] Oleksandr Sushko, Rostyslav Dubrovka, and Robert S. Donnan, “Terahertz spectral domain computational analysis of hydration shell of proteins with increasingly complex tertiary structure”, The Journal of Physical Chemistry, 2013.

[16] Reid CB, Reese G, Gibson AP, and Wallace VP, “Terahertz Time Domain Spectroscopy of Human Blood”, IEEE, 2013

[17] Simon Ebbinghaus, Seung Joong Kim, Matthias Heyden, Xin Yu, Martin Gruebele, David M. Leitner, and Martina Havenith, “Protein Sequence- and pH-Dependent Hydration Probed by Terahertz Spectroscopy”, J. Am. Chem. Soc. 2008, 130, 2374-2375

[18] Katarzyna M. Tych, Christopher D. Wood, Andrew D. Burnett, Arwen R. Pearson, A. Giles Davies, Edmund H. Linfield and John E. Cunningham, “Probing temperature- and solvent-dependent protein dynamics using terahertz time-domain spectroscopy”, Journal of Applied Crystallography, 2014

[19] Kiyoung Jeong, Yong-Min Hub, Sang-Hoon Kim, Yeonji Park, Joo-Hiuk Son, Seung Jae Oh, and Jin-Suck Suh, “Characterization of blood using terahertz waves”, Journal of Biomedical Optics, Oct. 2013.

[20] Andrea Markelz, Scott Whitmire, Jay Hillebrecht, and Robert Brige, “THz time domain spectroscopy of biomolecular conformational modes

[21] Changlei Wang, Jixian Gong, Qirong Xing, Yanfeng Li, Feng Liu, Xueming Zhao, Lu Chai, Chingyue Wang, Aleksei M. Zheltikov “Application of terahertz time-domain spectroscopy in intracellular metabolite detection”, Journal of Biophotonics, 2010

[22] Faustino Wahaia, Irmantas Kasalynas, Dalius Seliuta, Gediminas Molis, Andrzej Urbanowicz, Catia D. Carvalho Silva, Fatima Carneiro, Gintaras Valusis, Pedro L. Granja, “Study of paraffin-embedded colon cancer tissue using terahertz spectroscopy”, Journal of Molecular Structure, 2015.

[23] Meng, Kun; Chen, Tu-nan; Chen, Tao; Zhu, Li-guo; Liu, Qiao; Li, Zhao; Li, Fei; Zhong, Sen-cheng; Li, Ze-ren;Feng, Hua; Zhao, Jian-heng, “Terahertz pulsed

spectroscopy of paraffin-embedded brain glioma", Journal of Biomedical Optics, July 2014.

[24] Xiang Yang, Xiang Zhao, Ke Yang, Yueping Liu, Yu Liu, Weiling Fu, and Yang Luo, "Biomedical Applications of Terahertz Spectroscopy and Imaging", Trends in Biotechnology, Oct. 2016.

[25] Xu, J., Plaxco, K. and Allen, S. (2006). Probing the collective vibrational dynamics of a protein in liquid water by terahertz absorption spectroscopy. *Protein Science*, 15(5), pp.1175-1181.

[26] Xu, J., Plaxco, K. and Allen, S. (2006). Collective Dynamics of Lysozyme in Water: Terahertz Absorption Spectroscopy and Comparison with Theory. *The Journal of Physical Chemistry B*, 110(47), pp.24255-24259.

[27] A. G. Markelz, J. R. Knab, J. Y. Chen, and Y. He, "Protein dynamical transition in terahertz dielectric response," Chem. Phys. Lett., vol. 442, pp. 413–417, 2007.

[28] J.-Y. Chen, J. R. Knab, S. Ye, Y. He, and A. G. Markelz, "Terahertz dielectric assay of solution phase protein binding," Appl. Phys. Lett., vol. 90, pp. 243901-1–243901-3, 2007.

[29] Bahar, I., Lezon, T., Bakan, A. and Shrivastava, I. (2010). Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. *Chemical Reviews*, 110(3), pp.1463-1497

[30] Andrea G. Markelz, "Terahertz Dielectric Sensitivity to Biomolecular Structure and Function", IEEE Journal of Selected Topics in Quantum Electronics, 2008.

[31] Changlei Wang, Jixian Gong, Qirong Xing, Yanfeng Li, Feng Liu, Xueming Zhao, Lu Chai, Chingyue Wang, Aleksei M. Zheltikov "Application of terahertz time-

domain spectroscopy in intracellular metabolite detection", Journal of Biophotonics, 2010.

[32] Dibo Hou, Xian Li, Jinhui Cai, Yehao Ma, Xusheng Kang, Pingjie Huang, and Guangxin Zhang, "Terahertz spectroscopic investigation of human gastric normal and tumor tissues", IOP Science, 2016

[33] Thomas Elsaesser, Klaus Reimann, and Michael Woernerb, "Focus: Phase-resolved nonlinear terahertz spectroscopy—From charge dynamics in solids to molecular excitations in liquids", The Journal of Chemical Physics, 2015

[34] Ian A. Finneran, Ralph Welsch, Marco A. Allodia, Thomas F. Miller III, and Geoffrey A. Blakea, "Coherent two-dimensional terahertz-terahertz-Raman spectroscopy", PNAS, 2016

[35] Peter Hamm, and Janne Savolainen, "Two-dimensional-Raman-terahertz spectroscopy of water: Theory", The Journal of Chemical Physics, 2012

[36] Janne Savolainen, Saima Ahmed, and Peter Hamm, "Two-dimensional Raman-terahertz spectroscopy of water", PNAS, 2013

[37] Sushko, O., Dubrovka, R. and Donnan, R. "Sub-terahertz spectroscopy reveals that proteins influence the properties of water at greater distances than previously detected" [J. Chem. Phys. 142, 055101 (2015).

[38] R. Singh, D. George, C. Bae, K. Niessen and A. Markelz, "Modulated orientation-sensitive terahertz spectroscopy", *Photonics Research*, vol. 4, no. 3, p. A1, 2016.

[39] C. Mattos, "Protein–water interactions in a dynamic world", *Trends in Biochemical Sciences*, vol. 27, no. 4, pp. 203-208, 2002.

[40] M. Woerner, W. Kuehn, P. Bowlan, K. Reimann and T. Elsaesser, "Ultrafast two-dimensional terahertz spectroscopy of elementary excitations in solids", *New Journal of Physics*, vol. 15, no. 2, p. 025039, 2013.

[41] Jing Xu, Jhenny Galan, Gerry Ramian, Pavlos Savvidis, Anthony Scopatz, Robert R. Birge, S. James Allen, Kevin Plaxco, "Terahertz Circular Dichroism Spectroscopy of Biomolecules", Nov. 2003

[42] Choi, W.J., Cheng, G., Huang, Z. *et al.* Terahertz circular dichroism spectroscopy of biomaterials enabled by kirigami polarization modulators. *Nat. Mater.* **18,** 820–826 (2019).

[43] Sun, B. and Yu, Y. (2018). Analysis of circular dichroism in chiral metamaterial at terahertz frequencies. *Journal of Physics D: Applied Physics*, 52(2), p.025105.

[44] X. L. Tian, Y. D. Gong and H. Dong, "Simulation on circular dichroism of the water clusters in THz range," *35th International Conference on Infrared, Millimeter, and Terahertz Waves*, Rome, 2010, pp. 1-1.

[45] S. Zhang et al., "Photoinduced handedness switching in terahertz chiral metamolecules", *Nature Communications*, vol. 3, no. 1, 2012.

[46] Hinchliffe, A. and Munn, R. (1985). *Molecular electromagnetism*. Chichester: J. Wiley & Sons.

[47] R. Lewis, *Terahertz physics*. Cambridge: Cambridge University Press, 2012

[48] Philip Ball, "A handle on handness", nature, 22 June 2000, https://www.nature.com/news/2000/000622/full/news000622-10.html

[49] S. Kelly and N. Price, "The Use of Circular Dichroism in the Investigation of Protein Structure and Function", *Current Protein & Peptide Science*, vol. 1, no. 4, pp. 349-384, 2000.

[50] Anshuman Kumar, Reinhard Schweitzer-Stenner, and Bryan M. Wong, "A New Interpretation of the Structure and Solvent Dependence of the far UV circular Dichroism Spectrum of Short Oligopeptide", *Cehm. Commun., 2019, 55, 5701.*

[51] Böhm, G., Muhr, R., & Jaenicke, R. (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *"Protein Engineering, Design and Selection"*, *5*(3), 191–195.

[52] Nagy, G., Igaev, M., Jones, N. C., Hoffmann, S. V., & Grubmüller, H. (2019). SESCA: Predicting Circular Dichroism Spectra from Protein Molecular Structures. *Journal of Chemical Theory and Computation*, *15*(9), 5087–5102.

[53] Y. Wei, A. A. Thyparambil, and R. A. Latour, "Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230nm," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1844, no. 12, pp. 2331–2337, 2014.

[54] N. J. Greenfield, "Using circular dichroism spectra to estimate protein secondary structure," *Nature Protocols*, vol. 1, no. 6, pp. 2876–2890, 2006.

[55] Kelly, S. and Price, N. (2000), "The Use of Circular Dichroism in the Investigation of Protein Structure and Function", *Current Protein & Peptide Science*, 1(4), pp.349-384

[56] S. Kelly, T. Jess and N. Price, "How to study proteins by circular dichroism", *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1751, no. 2, pp. 119-139, 2005.

[57] B. Wallace, "Using Circular Dichroism (CD) and Synchrotron Radiation Circular Dichroism (SRCD) Spectroscopy to Study Membrane Proteins", *Biophysical Journal*, vol. 98, no. 3, pp. 209a-210a, 2010.

[58] D. Kurouski, "Advances of Vibrational Circular Dichroism (VCD) in bioanalytical chemistry. A review", *Analytica Chimica Acta*, vol. 990, pp. 54-66, 2017.

[59] G. Magyarfalvi, G. Tarczay and E. Vass, "Vibrational circular dichroism", *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, no. 3, pp. 403-425, 2011.

[60] Choi, W.J., Cheng, G., Huang, Z. *et al.* Terahertz circular dichroism spectroscopy of biomaterials enabled by kirigami polarization modulators. *Nat. Mater.* **18,** 820–826 (2019).

[61] Sun, B. and Yu, Y. (2018). Analysis of circular dichroism in chiral metamaterial at terahertz frequencies. *Journal of Physics D: Applied Physics*, 52(2), p.025105.

[62] Jing Xu, Jhenny Galan, Gerry Ramian, Pavlos Savvidis, Anthony Scopatz, Robert R. Birge, S. James Allen, Kevin Plaxco, “Terahertz Circular Dichroism Spectroscopy of Biomolecules”, Nov. 2003.

[63] P. Polavarapu and C. Zhao, "Vibrational circular dichroism: a new spectroscopic tool for biomolecular structural determination", *Fresenius' Journal of Analytical Chemistry*, vol. 366, no. 6-7, pp. 727-734, 2000.

[64] J. Choi and M. Cho, "Terahertz Chiroptical Spectroscopy of an α-Helical Polypeptide: A Molecular Dynamics Simulation Study", *The Journal of Physical Chemistry B*, vol. 118, no. 45, pp. 12837-12843, 2014.

[65] Kishore Padmaraju, "Faraday Rotation", Department of Physics and Astronomy, University of Rochester, Rochester, https://www.google.com/search?q=Kishore+Padmaraju%2C+%E2%80%9CFaraday+Rotation%E2%80%9D&oq=Kishore+Padmaraju%2C+%E2%80%9CFaraday+Rotation%E2%80%9D&aqs=chrome..69i57.2596j0j4&sourceid=chrome&ie=UTF-8

[66] Yang, B; Wylde, RJ; Martin, DH; Goy, P; Donnan, RS; Caroopen, S, "Determination of the Gyrotropic Characteristics of Hexaferrite Ceramics From 75 to 600 GHz",

[67] G. Magyarfalvi, G. Tarczay and E. Vass, "Vibrational circular dichroism", *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, no. 3, pp. 403-425, 2011.

[68] J. Xu, K. W. Plaxco, and S. J. Allen, "Probing the collective vibrational dynamics of a protein in liquid water by terahertz absorption spectroscopy," protein Science, vol. 15, no. 5, pp. 1175–1181, 2006.

[69] O. Sushko, "Terahertz dielectric study of bio-molecules using time-domain spectrometry and molecular dynamics simulations," Ph.D. dissertation, Sch. Electron. Eng. Comput. Sci., Queen Mary Univ. of London, London, U.K., 2014.

[70] The web pages of Thomas Keating Ltd., Billingshurst, UK.

[71] B. Yang and R. Donnan, "Enhanced rapid and accurate sub-THz magneto-optical characterization of hexaferrite ceramics", Journal of Magnetism and Magnetic Materials, vol. 323, no. 15, pp. 1992-1997, 2011.

[72] W. Withayachumnankul and M. Naftaly, "Fundamentals of Measurement in Terahertz Time-Domain Spectroscopy", *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 35, no. 8, pp. 610-637, 2013.

[73] R. Lewis, *Terahertz physics*. Cambridge: Cambridge University Press, 2012.

[74] P. H. Siegel, "Terahertz Pioneer: David H. Austin", IEEE Transactions on Terahertz Science and Technology, Vol. 1, No. 1, September 2011.

[75] Hangyo, M., Tani, M. and Nagashima, T. (2005). Terahertz Time-Domain Spectroscopy of Solids: A Review. International Journal of Infrared and Millimeter Waves, 26(12), pp.1661-1690.

[76] Acbas, G., Niessen, K., Snell, E. and Markelz, A. (2014). Optical measurements of long-range protein vibrations. Nature Communications, 5(1).

[77] R. J. Falconer and A. G. Markelz, "Terahertz Spectroscopic Analysis of Peptides and Proteins," Journal of Infrared, Millimeter, and Terahertz Waves, vol. 33, no. 10, pp. 973–988, 2012.

[78] Andrea Markelz, Scott Whitmire, Jay Hillebrecht, and Robert Brige, "THz time domain spectroscopy of biomolecular conformational modes", Phys. Med. Biol. 47 (2002) 3797-3805

[79] M.M. Nazarov, A.P. Shkurinov, E.A. Kuleshov, V.V. Tuchin, "Terahertz time-domain spectroscopy pf biological tissues", Quantum Electronics, 2008.

[80] Trung Quan Luong, Pramod Kumar Verma, Rajib Kumar Mitra, and Martina Havenith, “Do Hydration Dynamics Follow the Structural Perturbation during Thermal Denaturation of a Protein: A Terahertz Absorption Study”, Biophysical Journal, 2011.

[81] Zhang, X.-C and J. Xu, “Introduction to THz wave photonics”, Springer, 2010, vol. 29.

[82] J. Klier et al., "Influence of Substrate Material on Radiation Characteristics of THz Photoconductive Emitters", *International Journal of Antennas and Propagation*, vol. 2015, pp. 1-7, 2015.

[83] Wu, Q. and Zhang, X., “Free-space electro-optics sampling of mid-infrared pulses”. Applied Physics Letters, 71(10), 1997, pp.1285-1286.

[84] Ibrahim, A., Férachou, D., Sharma, G., Singh, K., Kirouac-Turmel, M. and Ozaki, T., “Ultra-high dynamic range electro-optic sampling for detecting millimeter and sub-millimeter radiation”. Scientific Reports, 6(1), 2016.

[85] Yang Zeng, “A Fundamental Critical Assessment of Efficiencies in Terahertz Time-Domain Spectroscopy Systems”, Ph.D Thesis for Queen Mary University of London, 2017.

[86] Y.-S. Lee, *Principles of terahertz science and technology*, 1st ed. New York, Springer, 2008

[87] Theuer, M., “Terahertz Time Domain Spectroscopy Systems For Fundamental And Industrial Applications”, 2009, Göttingen: Cuvillier

[88] J. Xu, T. Yuan, S. Mickan and X. Zhang, "Limit of Spectral Resolution in Terahertz Time-Domain Spectroscopy", *Chinese Physics Letters*, vol. 20, no. 8, pp. 1266-1268, 2003.

[89] R. Singh, D. George, C. Bae, K. Niessen and A. Markelz, "Modulated orientation-sensitive terahertz spectroscopy", *Photonics Research*, vol. 4, no. 3, p. A1, 2016.

[90] C. Mattos, "Protein–water interactions in a dynamic world", *Trends in Biochemical Sciences*, vol. 27, no. 4, pp. 203-208, 2002

[91] J. Choi and M. Cho, "Terahertz Chiroptical Spectroscopy of an α-Helical Polypeptide: A Molecular Dynamics Simulation Study", *The Journal of Physical Chemistry B*, vol. 118, no. 45, pp. 12837-12843, 2014.

[92] Laman, N., Harsha, S., Grischkowsky, D. and Melinger, J. (2008). High-Resolution Waveguide THz Spectroscopy of Biological Molecules. *Biophysical Journal*, 94(3), pp.1010-1020.

[93] C.R. Nave, "Interaction of radiation with Matter", Georgia State University, *Hyperphysics.phy-astr.gsu.edu*, 2016.

[94] R. Feldberg, "Structure and Function in Lysozyme: Tufts University", [Online]. https://ase.tufts.edu/biology/molecvisual/bio152/rightlyso.html.

[95] "Bovine Serum Albumin", https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Product_Information_Sheet/a2153pis.pdf, Sigma-Aldrich.

[96] I. Kamphuis and J. Drenth, "Structure of Papain Refined at 1.65 Angstroms Resolution", RCSB-PDB, 1986, https://www.rcsb.org/structure/9PAP

[97] Zhang, C. and Durbin, S. (2006). Hydration-Induced Far-Infrared Absorption Increase in Myoglobin. *The Journal of Physical Chemistry B*, 110(46), pp.23607-23613.

[98] S. Saha, "Application of terahertz spectroscopy to the characterization of biological samples using birefringence silicon grating", *Journal of Biomedical Optics*, vol. 17, no. 6, p. 067006, 2012. Available: 10.1117/1.jbo.17.6.067006.

[99] B. Rašković, M. Popović, S. Ostojić, B. Anđelković, V. Tešević and N. Polović, "Fourier transform infrared spectroscopy provides an evidence of papain denaturation and aggregation during cold storage", *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 150, pp. 238-246, 2015.

[100] M. Holyavka et al., "Influence of UV radiation on molecular structure and catalytic activity of free and immobilized bromelain, ficin and papain", *Journal of Photochemistry and Photobiology B: Biology*, vol. 201, p. 111681, 2019.

[101] O. Sushko, K. Shala, R. Dubrovka and R. Donnan, "Revised metrology for enhanced accuracy in complex optical constant determination by THz-time-domain spectrometry", *Journal of the Optical Society of America A*, vol. 30, no. 5, p. 979, 2013

[102] A. Charkhesht, C. Regmi, K. Mitchell-Koch, S. Cheng and N. Vinh, "High-Precision Megahertz-to-Terahertz Dielectric Spectroscopy of Protein Collective Motions and Hydration Dynamics", *The Journal of Physical Chemistry B*, vol. 122, no. 24, pp. 6341-6350, 2018.

[103] F. Lipps, S. Levy and A. Markelz, "Hydration and temperature interdependence of protein picosecond dynamics", *Physical Chemistry Chemical Physics*, vol. 14, no. 18, p. 6375, 2012.

[104] N. Penkov, V. Yashin, E. Fesenko, A. Manokhin and E. Fesenko, "A Study of the Effect of a Protein on the Structure of Water in Solution Using Terahertz Time-Domain Spectroscopy", *Applied Spectroscopy*, vol. 72, no. 2, pp. 257-267, 2017.

[105] F. Bibi, M. Villain, C. Guillaume, B. Sorli and N. Gontard, "A Review: Origins of the Dielectric Properties of Proteins and Potential Development as Bio-Sensors", *Sensors*, vol. 16, no. 8, p. 1232, 2016.

[106] T. Simonson and D. Perahia, "Dielectric properties of proteins from simulations: tools and techniques", *Computer Physics Communications*, vol. 91, no. 1-3, pp. 291-303, 1995.

[107] F. Novelli et al., "Time-Domain THz Spectroscopy Reveals Coupled Protein–Hydration Dielectric Response in Solutions of Native and Fibrils of Human Lysozyme", *The Journal of Physical Chemistry B*, vol. 121, no. 18, pp. 4810-4816, 2017.

[108] J. Cheng, R. Jones, O. Sushko, Y. Tashiro and R. Donnan, "Quasi-Optical Sub-THz Circular Dichroism Spectroscopy of solvated Myoglobin", *IEEE Transactions on Terahertz Science and Technology*, pp. 1-1, 2020.

[109] A. Bonincontro and G. Risuleo, "Dielectric spectroscopy as a probe for the investigation of conformational properties of proteins", *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 59, no. 12, pp. 2677-2684, 2003.

[110] U. Kaatze, "Complex permittivity of water as a function of frequency and temperature", *Journal of Chemical & Engineering Data*, vol. 34, no. 4, pp. 371-374, 1989.

[111] H. Yada, M. Nagai and K. Tanaka, "The intermolecular stretching vibration mode in water isotopes investigated with broadband terahertz time-domain spectroscopy", *Chemical Physics Letters*, vol. 473, no. 4-6, pp. 279-283, 2009.

[112] G. Papadakis, D. Fleischman, A. Davoyan, P. Yeh and H. Atwater, "Optical magnetism in planar metamaterial heterostructures", *Nature Communications*, vol. 9, no. 1, 2018.

[113] I. Pupeza, R. Wilk and M. Koch, "Highly accurate optical material parameter determination with THz time-domain spectroscopy", *Optics Express*, vol. 15, no. 7, p. 4335, 2007.

[114] G. Wilmink et al., "Development of a compact terahertz time-domain spectrometer for the measurement of the optical properties of biological tissues", *Journal of Biomedical Optics*, vol. 16, no. 4, p. 047006, 2011.

[115] C. Ronne, L. Thrane, P. Åstrand, A. Wallqvist, K. Mikkelsen and S. Keiding, "Investigation of the temperature dependence of dielectric relaxation in liquid water by THz reflection spectroscopy and molecular dynamics simulation", *The Journal of Chemical Physics*, vol. 107, no. 14, pp. 5319-5331, 1997.

[116] A. Fitzgerald, E. Pickwell-MacPherson and V. Wallace, "Use of Finite Difference Time Domain Simulations and Debye Theory for Modelling the Terahertz Reflection Response of Normal and Tumour Breast Tissue", *PLoS ONE*, vol. 9, no. 7, p. e99291, 2014.

[117] “Protein Dipole Moment Server”, https://dipole.weizmann.ac.il/dipol/

[118] O. Cherkasova, M. Nazarov and A. Shkurinov, "Properties of aqueous solutions in THz frequency range", *Journal of Physics: Conference Series*, vol. 793, p. 012005, 2017.

[119] O. Smart, "Overview of molecular forces: Non-bonded Interactions", 1996. [Online].http://www.cryst.bbk.ac.uk/PPS2/course/section7/os_non.html#non_hbond

[120] S. V. Frank, E. Leiss-Holzinger, M. Pfleger, and C. Rankl, “Terahertz Time-Domain Polarimetry in Reflection for Film Characterization,” *Sensors*, vol. 20, no. 12, p. 3352, 2020.

[121] J. Qin, L. Xie, and Y. Ying, “Rapid analysis of tetracycline hydrochloride solution by attenuated total reflection terahertz time-domain spectroscopy,” *Food Chemistry*, vol. 224, pp. 262–269, Jun. 2017.

[122] M. NAGAI and K. TANAKA, “Terahertz Time-Domain Attenuated Total Reflection Spectroscopy in Amino-Acid Solution,” *The Review of Laser Engineering*, vol. 33, no. 12, pp. 843–847, 2005.